10/27/95 18C1

PATENT
Attorney Docket No. 3495.0010-19

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:

Marc ALIZON et al.

Serial No.: 08/308,218

Filed: September 19, 1994

For: DNA SEQUENCES AND PEPTIDES OF HUMAN IMMUNODEFICIENCY VIRUS (HIV-1)

Group Art Unit: 1805 ~~1804~~

Examiner: J. Railey

RECEIVED
OCT 17 1995
GROUP 1800

TRANSMITTAL LETTER

Assistant Commissioner for Patents
Washington, D.C. 20231

Sir:

Enclosed is a response to the Office Action of March 27, 1995. The item checked below are appropriate:

[X] Applicants hereby petition for a three month extension of time to respond to the above Office Action. The fee of $870.00 for the Extension is enclosed.

The claims are calculated below:

	Claims Remaining After Amendment		Highest Number Previously Paid	Present Extra	Rate	Additional Fee
Total	1	-	2		x $ 22	$
Indep.	1	-	1		x $ 74	
[ ] First Presentation of Multiple Dep. Claim(s)					+ $230	
					Subtotal	$
					Reduction by ½ if small entity	-
					TOTAL	$ -0-

[ ] A fee of $_____ to cover the cost of the additional claims added by this response is enclosed.

[ ] A fee of $_____ to cover _____ is enclosed.

[X] A check for $870.00 to cover the above fee is enclosed.

If there are any other fees due in connection with the filing of this response, please charge the fees to our Deposit Account No. 06-0916. If a fee is required for an extension of time under 37 C.F.R. § 1.136 not

accounted for above, such an extension is requested and the fee should also be charged to our Deposit Account.

Date: September 25, 1995

By: _____
David J. Kulik
Registration No. 36,576
FINNEGAN, HENDERSON, FARABOW,
GARRETT & DUNNER
1300 I Street, N.W.
Washington, D.C. 20005-3315
(202) 408-4000

870117 180f

PATENT

Attorney Docket No. 3495.0010-19

#21 NSM 11/12/95

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:

Marc ALIZON et al.

Serial No.: 08/308,218 — Group Art Unit: 1804

Filed: September 19, 1994 — Examiner: Railey, J.

For: DNA SEQUENCES AND PEPTIDES OF HUMAN IMMUNODEFICIENCY VIRUS (HIV-1)

Assistant Commissioner for Patents
Washington, D.C. 20231

Sir:

RESPONSE TO PAPER NO. 20

In response to the Office action mailed March 27, 1995, the time for response to which is extended with the accompanying transmittal letter inlcuding a petition, reconsideration and reexamination are respectfully requested.

One rejection stands against claim 13 of this application. Claim 13 is rejected under 35 U.S.C. § 112, first paragraph, for allegedly failing to provide an adequate written description of the invention and allegedly failing to provide an enabling disclosure. This rejection is respectfully traversed.

Applicants response separates this rejection into the two issues of first fulfilling the written description requirement and second fulfilling the enablement requirement.

LAW OFFICES
FINNEGAN, HENDERSON,
FARABOW, GARRETT
& DUNNER, L.L.P.
1300 I STREET, N. W.
WASHINGTON, DC 20005
202-408-4000

Applicants' specification literally sets forth the nucleic acid claimed. As noted by the Examiner, the last line of page 12 of the specification indicates that the claimed nucleic acid is specifically disclosed. Paper No. 20 at page 2. To find a literal disclosure of the subject matter claimed, one could read the sequence on Figs. 11 and 12. With this literal support, it is not understood how there can be a question as to the specifications' fulfillment of the written description requirement.

The purpose of the written description of § 112 has been clearly stated by the Federal Circuit:

> The purpose of the "written description" requirement is broader than to merely explain how to "make and use"; the applicant must also convey with reasonable clarity to those skilled in the art that, as of the filing date sought, he or she was in possession *of the invention*. The invention is, for purposes of the "written description" inquiry, *whatever is now claimed*. (emphasis in original)

Vas-Cath Inc. vs. Mahurkar, 935 F.2d 1555, 1563-1564, 19 U.S.P.Q.2d 1111, 1117 (Fed Cir. 1991). In this case, by specifically setting out in the specification the nucleic acid as it is now claimed, one of ordinary skill in the art would certainly have known that applicants' were in possession of the claimed invention. Therefore, applicants submit that they have met the written description requirement of 35 U.S.C. § 112, first paragraph.

As noted by the Examiner (Paper No. 20 at pages 2-3), the specification also sets forth specific uses for the

LAW OFFICES
FINNEGAN, HENDERSON, FARABOW, GARRETT & DUNNER, L.L.P.
1300 I STREET, N. W.
WASHINGTON, DC 20005
202-408-4000

claimed invention. Other uses, such as in cloning HIV-1 DNA and related DNA, would be readily apparent to one of skill in the art.

The Examiner asserts, however, that the specification does not set forth how to make and use the invention for the asserted purposes. Paper No. 20 at page 3. The recited sequence could have been made by chemical synthesis or recombinant DNA methods using applicants' disclosure. No barriers exist, and none have been asserted, to making the claimed invention.

With respect to the use of the claimed invention, applicants note that "use" has been clearly defined by the Federal Circuit in a case such as this. "Usefulness in patent law, and in particular in the context of pharmaceutical inventions, necessarily includes the expectation of further research and development." In re Brana, 51 F.3d 1560, 1568, 34 U.S.P.Q.2d 1437, 1442 (Fed. Cir. 1995). Therefore, applicants need not demonstrate that the claimed invention can be used to produce a marketable or finished diagnostic product, vaccine, or therapeutic.

Furthermore, the Examiner has supplied no reasons why the claimed invention could not be used to clone HIV-1 sequences or as a hybridization probe. Since only one use suffices and each of these uses would be readily apparent to one of skill in the art, this rejection, *inter alia*, is in error.

LAW OFFICES
FINNEGAN, HENDERSON, FARABOW, GARRETT & DUNNER, L.L.P.
1300 I STREET, N. W.
WASHINGTON, DC 20005
202-408-4000

The Examiner suggests that the known and routine methods for expressing an open reading frame would not apply to the claimed invention. Brown et al. is cited as alleged evidence for what "must be considered for successful expression of a given gene sequence." Paper No. 20 at page 3.

However, the skilled artisan was often faced with the routine task of expressing an open reading frame. Exhibits 1 and 2, attached hereto, show general methods for identifying and expressing open reading frames and localizing important immunoreactive regions of polypeptides expressed from the open reading frames. Gray et al., Exhibit 1, shows vectors and methods for expressing open reading frames known to one of skill in the art at the time of filing. Nunberg et al., Exhibit 2, shows methods for expressing and determining antigenic peptides from an open reading frame. In addition, as commercial enterprises were, at the time of filing, producing and selling vectors for expressing and analyzing open reading frames, applicants submit that the Examiner's assertions are incorrect. One did not need specific instructions on how to express the sequence as claimed.

The Examiner also mentions the glycosylation of the expressed polypeptide. Paper No. 20 at page 3. Since this plainly is the type of "further research and development" discussed in In re Brana (see above), these arguments do not appear relevant to the patentable use of the claimed invention. The arguments addressing useful vaccines and

LAW OFFICES
FINNEGAN, HENDERSON,
FARABOW, GARRETT
& DUNNER, L.L.P.
1300 I STREET, N. W.
WASHINGTON, DC 20005
202-408-4000

diagnostics at page 4 of Paper No. 20 also encompass "further research and development" that applicants need not show for the claimed invention.

With respect to the use of the claimed invention as a hybridization probe, the Examiner asserts "no specific conditions or methods are given which would allow distinction between HIV-1 and other retroviruses." Paper No. 20 at page 5. However, why such a demonstration need be shown is not understood.

One of skill in the art was clearly familiar with hybridization methods and assays as well as with the possible factors that can be varied. Well known also was the formula for determining the melting temperature of a "hybrid." See the Maniatis excerpt (Maniatis et al. Molecular Cloning, 1982, pp. 324-5, 388-9), attached as Exhibit 3, showing the routine considerations for hybridization assays known to one of skill in the art. See also the Wu excerpt (Wu et al. eds., Methods in Enzymology, 1983, vol. 100, pp. 266-285), attached as Exhibit 4. Applicants set forth the entire HIV-1 DNA sequence with the G+C content available from that sequence. As described in both Maniatis and Wu, one could have routinely calculated the melting temperature, thus the hybridization conditions, for assaying a sample with HIV-1 DNA with this information. No further guidance is needed.

Applicants contend that one of skill in the art could have easily made a probe that only detects HIV-1 from the

LAW OFFICES
FINNEGAN, HENDERSON, FARABOW, GARRETT & DUNNER, L.L.P.
1300 I STREET, N. W.
WASHINGTON, DC 20005
202-408-4000

present disclosure. However, even if the alleged cross-hybridization to other HTLV viruses the Examiner mentions (page 5 of Paper No. 20) did occur, under the patent laws, this would not prevent the use of the claimed invention. A probe that identifies all the members of the HTLV family can be a useful probe for that purpose.

In fact, the Examiner's contentions regarding the cross-hybridization cannot be properly based upon the contents of the Arya et al. and Hahn et al. documents. Both Arya et al., at page 928 and the legend of Fig. 2, and Hahn et al., at page 168 and the legend of Fig. 4., express the desire to show "homology" and "relatedness" between the members of the HTLV family of viruses. Hahn et al. explicitly states "low stringency conditions" in the legend of Fig. 4, which one of skill in the art knows would foster cross-hybridization.

However, the Examiner uses Arya et al. and Hahn et al. to allegedly demonstrate that they show something beyond their contents, i.e. that a hybridization assay discriminating between the members of the HTLV family is not possible with the claimed invention. This is especially inappropriate in a case such as this where the objective of both Arya et al. and Hahn et al. are obviously different from the use the Examiner is analyzing for the claimed invention.

For these reasons, applicants claimed invention is enabled by the specification and a *prima facie* case of lack of enablement has not been set forth. Applicants

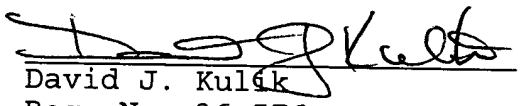
LAW OFFICES
FINNEGAN, HENDERSON,
FARABOW, GARRETT
& DUNNER, L.L.P.
1300 I STREET, N. W.
WASHINGTON, DC 20005
202-408-4000

respectfully request withdrawal of this objection and rejection.

If there are any other fees due in connection with the filing of this Response, please charge such fees to our Deposit Account No. 06-0916. If a fee is required for an Extension of Time under 37 C.F.R. § 1.136 not accounted for above, such an Extension is requested and the fee should also be charged to our Deposit Account.

Respectfully submitted,

FINNEGAN, HENDERSON, FARABOW, GARRETT & DUNNER

By: 
David J. Kulik
Reg. No. 36,576

Date: September 25, 1995

LAW OFFICES
FINNEGAN, HENDERSON, FARABOW, GARRETT & DUNNER, L.L.P.
1300 I STREET, N. W.
WASHINGTON, DC 20005
202-408-4000

Exhibit 1

Proc. Natl. Acad. Sci. USA
Vol. 79, pp. 6598–6602, November 1982
Genetics

Open reading frame cloning: Identification, cloning, and expression of open reading frame DNA

(*lacZ* gene fusion/frameshift mutant/hybrid protein)

MARK R. GRAY*, HILDUR V. COLOT*, LEONARD GUARENTE†, AND MICHAEL ROSBASH*

*Department of Biology, Brandeis University, Waltham, Massachusetts 02254; and †Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Communicated by Norman Davidson, July 14, 1982

ABSTRACT A plasmid was constructed that facilitates the cloning and expression of open reading frame DNA. A DNA fragment containing a bacterial promoter and the amino terminus of the *cI* gene of bacteriophage λ was fused to an amino-terminally deleted version of the *lacZ* gene. An appropriate cloning site was inserted between these two fragments such that a frameshift mutation was introduced upstream of the lacZ-encoding DNA. This cloning vehicle produces a relatively low level of β-galactosidase activity when introduced into *Escherichia coli*. The insertion of foreign DNA at the cloning site can reverse the frameshift mutation and generate plasmids that produce a relatively high level of β-galactosidase activity. A large fraction of these plasmids produce a fusion protein that has a portion of the λ cI protein at the amino terminus, the foreign protein segment in the middle, and the lacZ polypeptide at the carboxyl terminus. The production of a high level of β-galactosidase and a large fusion polypeptide guarantees the cloning of a DNA fragment with at least one open reading frame that traverses the entirety of the fragment. Hence, the method can identify, clone, and express (as part of a larger fusion polypeptide) open reading frame DNA from among a large collection of DNA fragments.

The use of recombinant DNA technology has increased our understanding of the organization and expression of eukaryotic DNA. While many goals can be achieved with existing technology, certain features of eukaryotic gene organization are still attainable only with difficulty. In particular, the identification of protein-coding segments within a large block of eukaryotic DNA is quite challenging. One should be able to take advantage of properties that are unique or highly enriched in coding DNA. Perhaps the most ubiquitous and specific feature of coding DNA is the presence of open reading frames. Indeed, analysis at the DNA sequence level has proved invaluable in identifying genes or gene segments. Although there are many examples of genes in which relatively small exons are interrupted by relatively large and numerous introns, this characteristic appears to be relatively infrequent in lower eukaryotes. Also, higher eukaryotic genes have been analyzed that contain open reading frames of substantial length.

In this report, we describe a method for cloning pieces of DNA based on this property. We have exploited the elegant work of Beckwith and co-workers (1), who have demonstrated the power and utility of gene and operon fusions to the *Escherichia coli* β-galactosidase gene (the *lacZ* gene). Many of these studies rely on the fact that the enzyme, β-galactosidase (β-D-galactoside galactohydrolase, EC 3.2.1.23) is often biologically active when it carries other protein segments at its amino terminus. Indeed, it appears that in yeast, as well as in *E. coli*, β-galactosidase can carry a wide range of protein sequences at its amino terminus and still retain substantial biological activity (2, 3). We have taken advantage of this property to create a *lacZ* vector for the cloning of open reading frames. We have introduced a frameshift mutation by inserting a DNA linker at the junction of a *cIlacZ* fused gene. As expected, this frameshift results in a relatively low level of *lacZ* activity. The DNA linker encodes a cloning site. Consequently, the vector is able to screen for segments of open reading frame DNA, since a continuous open reading frame sequence is a necessary condition for high-level expression of the *lacZ* sequence downstream from the insert. A successful construction results in the synthesis of a lacZ fusion polypeptide that includes the protein segment coded for by the inserted open reading frame. The vector is able to identify, clone, and express open reading frame DNA in one step.

METHODS AND MATERIALS

Materials. Restriction enzymes, T4 DNA ligase, and BAL-31 nuclease were purchased from New England BioLabs and Bethesda Research Laboratories. Calf intestine phosphatase was purchased from Boehringer-Mannheim; 5-bromo-4-chloro-3-indolyl β-D-galactoside (XGal), *o*-nitrophenyl β-D-galactopyranoside (ONPG), and isopropyl β-D-thiogalactoside were purchased from Sigma. XGal plates and MacConkey agar plates (Difco) were prepared as described by Miller (4). *Sma* I/*Bam*HI adapter was purchased from New England BioLabs.

Antibodies. Rabbit anti-β-galactosidase was a gift of Beth Rasmussen, rabbit anti-cI was a gift of Jeffrey Roberts, and the ^{125}I-labeled IgG fraction of a goat anti-rabbit immunoglobulin Fc serum was a gift of Susan Lowey and Joan Press.

Bacterial Strains. The lac deletion strain LG90 (F^- Δ*lac pro* XIII) was used exclusively (2).

Plasmid DNA Preparation. Plasmid DNA was prepared by standard CsCl/ethidium bromide centrifugation methods (5) or from minilysates prepared by using the alkaline lysis procedure (6).

Plasmid Constructions. Plasmid DNA was digested with commercial restriction enzymes as recommended by the suppliers. Where indicated, restricted DNA was treated with sufficient phosphatase [10 mM Tris·HCl (pH 8.0), 30 min at 37°C] to reduce self-ligation of a comparable amount of vector DNA by a factor of at least 1/100. DNA was sonicated in 5 ml of TE buffer (10 mM Tris·HCl, pH 8.3/1 mM EDTA)/0.2 M NaCl with six 30-sec bursts at maximum power, concentrated on a 0.2-ml column of DEAE-cellulose, eluted with TE buffer/1 M NaCl, and precipitated with ethanol. DNA was resuspended in TE buffer, digested lightly with BAL-31 nuclease (sufficient to

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviations: XGal, 5-bromo-4-chloro-3-indolyl β-D-galactoside; bp, base pair(s).

remove approximately 25 nucleotides from each end of a similar quantity of control restriction enzyme-digested DNA fragments) at 30°C, and fractionated by size on 10% acrylamide/Tris borate/EDTA (TBE) gels along with appropriate molecular weight markers (7).

Ligations. Cohesive-end ligations were carried out as follows: 100 ng of plasmid vector and various amounts of eukaryotic DNA were ligated for 3 hr at 22°C in a final volume of 100 μl with 0.1 (Bethesda Research Laboratories) unit of T4 DNA ligase. Blunt-end ligations were typically carried out as follows: 500 ng of plasmid vector and various amounts of potential insert DNA were ligated overnight at 16°C in a final volume of 10 μl with 1 unit of T4 DNA ligase.

Transformation. Bacterial transformation into LG90 was carried out by standard procedures (8), with the following modifications. All ligated DNA samples were phenol extracted and ethanol precipitated prior to transformation, which caused a reproducible 5- to 10-fold increase in transformation efficiency. Transformed cells (1 ml) were spread on the appropriate indicator plates containing ampicillin at 25 μg/ml. Colonies were scored for phenotype on MacConkey agar plates after 24 hr at 37°C. If the plates were crowded, the colony phenotype was scored at 48 hr.

Hybrid Protein Analysis. Three-milliliter samples of cells from a saturated culture (grown in L broth containing ampicillin at 50 μg/ml) were centrifuged, the pellet was suspended in 150 μl of 1.2-fold concentrated sample buffer (9) and heated at 100°C for 3 min, and the viscosity was reduced by repeated passages through a 22-gauge syringe needle. Between 4 and 8 μl of the protein samples were electrophoresed in 7.5% acrylamide gels (9). Protein blotting was carried out according to published procedures (10).

β-Galactosidase Assays. These were carried out according to Miller (4) except that cells were grown in L broth.

RESULTS

Our initial hypothesis was that a frameshift mutation placed near the amino terminus of an open reading frame sequence that contains the *lacZ* sequence at its carboxyl terminus would produce a plasmid that would be phenotypically lacZ⁻ when transformed into *E. coli*. Furthermore, the insertion of foreign DNA at or near the frameshift could correct the frameshift mutation and generate a *lacZ*⁺ insert DNA-containing plasmid. The two constructions shown in Fig. 1*A* were designed to test this hypothesis. A segment of the *cI* gene from bacteriophage λ terminating at a *Hin*dIII site at codon 157 was fused to a *lacIZ* fragment. The *cI* gene fragment also contains two *lac* promoters that are oriented toward *cI*. Both constructions were identical except that two different *lacZ*-containing plasmids were used. These two plasmids, pLG200 and pLG400, differ in DNA sequence only in the immediate vicinity of the *Hin*dIII site (12). Because the DNA sequence of the λ *cI* gene is also known (13), as well as the reading frame of the lacI–lacZ (lacIZ) fusion protein encoded by both pLG200 and pLG400 (12), we could predict the reading frame of the two resultant plasmids. Plasmid pMR1 has 24 bases between the lysine codon at amino acid 157 in the *cI* gene and the leucine codon at the beginning of the *lacI* fragment (Fig. 1*A*). Since none of these eight codons are stop codons, this construction reads in frame from the λ *cI* gene into the *lacIZ* fusion and consequently produces a high level of β-galactosidase activity. In contrast, pMR2, constructed from pLG200, has 10 bases between the AAG lysine codon of the λ *cI* gene and the CTG leucine codon of *lacI*. Because this number of intervening bases is not divisible by 3, this construction should shift the frame, resulting in improper reading of or termination in (or both) the *lacIZ* fusion sequence. This plasmid results in a relatively low level of β-galactosidase activity. In addition, the insertion of eukaryotic DNA into the *Hin*dIII site of pMR2 is able to generate colonies with a high level of lacZ activity (data not shown).

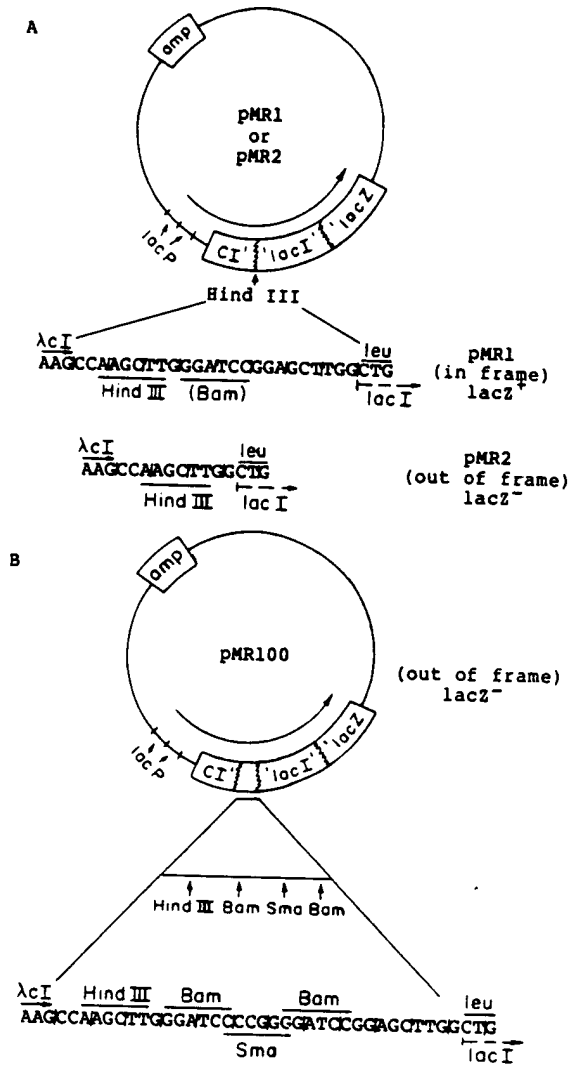


FIG. 1. Construction of vectors. (*A*) Plasmids pKB252 (11), pLG200, and pLG400 (12) were manipulated by standard procedures to produce the appropriate fragments for ligation. Briefly, pKB252 was digested with *Hae* III/*Hin*dIII and the largest fragment was purified. pLG200 (or pLG400) was digested with *Sma* I/*Hin*dIII and the largest fragment was purified. These two fragments were ligated together and ampicillin-resistant colonies were selected. The restriction sites of the resultant plasmids were verified. pMR1 was derived by using pLG400 and pMR2 was derived by using pLG200. Both plasmids, when transformed into LG90, form blue colonies on XGal plates. Thus, lac⁺ and lac⁻ refer to phenotypes on MacConkey agar plates. The DNA sequence at the joint is inferred from the sequence of the λ *cI* gene up to the *Hin*dIII site (13) and the sequence in the vicinity of the *Hin*dIII site of pLG200 and pLG400 (12). (*B*) One hundred nanograms of *Bam*HI-digested pMR1 and 2 ng of *Bam*HI/*Sma* I adapter were ligated (final vol, 100 μl) with 0.1 unit of ligase for 3 hr at room temperature. DNAs were transformed into LG90, and ampicillin-resistant lac⁻ colonies were scored on MacConkey agar plates. DNA sequences in the vicinity of the cloning sites were verified by direct sequence analysis.

To construct a vector of more utility, pMR1 was manipulated as in Fig. 1*B*. The construction inserted 10 base pairs (bp) at the *Bam*HI site and therefore changed the reading frame so that the *lacZ* sequence is no longer properly translated. Moreover,

it has created a blunt-ended *Sma* I site, flanked by two *Bam*HI sites, within this frameshifted region. The *Sma* I site provides a cloning site into which blunt-ended DNA can be cloned. Insert DNA can be excised by digestion with *Bam*HI.

To verify that this second-generation vector was suitable for cloning open reading frame DNA, *Hae* III-digested pBR322 was cloned into the *Sma* I site of pMR100. After transformation into LG90, 28 red colonies were chosen at random and analyzed for plasmid DNA insert size. All 28 clones yielded an additional *Bam*HI fragment when analyzed by acrylamide gel electrophoresis; 8 of these clones are shown in Fig. 2. A very limited subset of the possible pBR322 *Hae* III fragments are cloned when ampicillin-resistant colonies are chosen on the basis of a strong lacZ$^+$ phenotype. As expected, most white colonies from this ligation contain plasmid DNA with no detectable insert, since most of these colonies are presumably derived from recyclization of the vector (see Table 1). White colonies that harbor a plasmid with a DNA insert do not show the dramatic insert size preference seen with red colonies; DNA fragments of many different sizes are obtained from white colonies (data not shown).

The sequence of pMR100 predicts that two criteria must be met for a DNA fragment inserted into the *Sma I* site to generate a proper open reading frame between the λ *cI* fragment and the *lacIZ* fragment. (*i*) The inserted DNA must be of a size $3n - 1$ (where n is an integer) and (*ii*) the inserted DNA must contain no stop codons in the reading frame set by the frame of the λ *cI* gene. Examination of the size and sequence (in both orientations) of the pBR322 *Hae* III fragments predicts that, of the 22 *Hae* III fragments, only 3 meet the two criteria: a 104-bp fragment from the tetracycline-resistance gene (which contains the single pBR322 *Bam*HI site), an 89-bp fragment, and an 80-bp fragment, the latter two from the ampicillin-resistance gene. The 89-bp fragment has an appropriate open reading frame in both orientations, while the 104-bp fragment is cleaved by *Bam*HI into two fragments, a 78- and a 26-bp fragment. Consequently, the sequence data predict that 4 fragment orientations out of a possible 44 (22 fragments × two orientations) should produce red colonies. This value of 10% is consistent with experiment (Table 1). Moreover, there should be two size classes of insert fragments produced by *Bam*HI digestion, 103 and about 93 nucleotides (89- and 78- to 80-nucleotide fragments plus 14 nucleotides from the DNA between the *Bam*/*Sma* sites shown in Fig. 1). Consistent with expectation, 26 of the 28 red colonies examined at random have one of the two size classes of DNA insert shown in Fig. 2. Moreover, double digestion of plasmids derived from these red colonies confirms that some of the larger size DNA inserts are indeed derived from the 89-bp *Hae* III fragment, which has been cloned in both orientations.

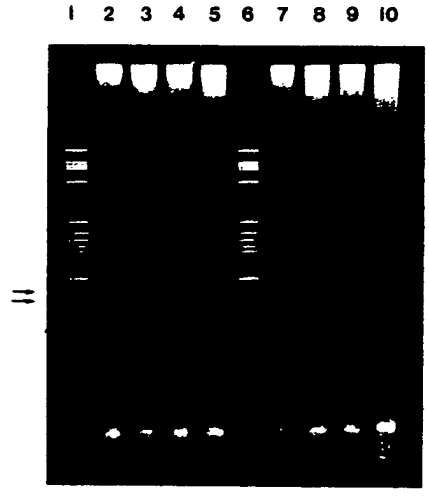


FIG. 2. Cloning of *Hae* III fragments of pBR322 into pMR100. pBR322 DNA was digested to completion with *Hae* III and the resultant fragments digested with phosphatase (to prevent multiple inserts). The phosphatase-digested DNA was ligated with pMR100 and transformed into LG90. Red colonies were picked, and DNA was isolated from minilysates and digested with *Bam*HI. DNA was analyzed on 10% acrylamide/Tris borate/EDTA gels. Lanes: 1 and 6, *Hae* III pBR322 fragment standards; 2–5 and 7–10, DNA from random red colonies. Arrows indicate the 104-bp and 89-bp *Hae* III pBR322 fragments.

Table 1. Cloning of pBR322 *Hae* III fragments into pMR100 and pMR200

Ligation mixture	Vector	Vector/pBR322 ratio*	Colonies	
			Red	White
1	pMR200†	1:1	349	13
2	pMR200	1:5	300	65
3	pMR100	1:5	11	594

Phosphatase-digested pBR322 *Hae* III fragments were ligated with *Sma* I-digested pMR100 or *Sma* I-digested pMR200. Ligated DNA was transformed into LG90 and colonies were scored on MacConkey agar plates. Because pMR200 is a lacZ$^+$ vector having the *Sma* I cloning site at the same position as pMR100, the fraction of white colonies ≈ the fraction of religated vector that contains an insert. This value was ≈20% in ligation mixture 2 but, in ligation mixture 3 under identical conditions, 11/605 colonies are red. Normalizing for the expected fraction of colonies from this ligation that contain an insert (≈20% ≈ same value as ligation mixture 2), the fraction of pBR322 *Hae* III fragments that can reverse the frameshift ≈ (11/605) × 0.2 ≈ 10%.

* Arbitrary.

† pMR200 was derived from pMR100 as a rare lacZ$^+$ colony (on MacConkey agar) after digestion of pMR100 with *Sma* I, self-ligation, and transformation. It is almost certainly a single-base-pair deletion of pMR100, placing the reading frame back in register. [The *Sma* I site (Fig. 1) is retained despite the deletion.] The size of the lacZ polypeptide is identical to that produced by pMR1 and all restriction sites in pMR100 are still present. Since almost all inserts into the *Sma* I site are lacZ$^-$, the vector can be used to measure insert frequency.

The above data suggest that the vector pMR100 can select suitable open reading frame fragments from among a larger group of DNA segments. Because of the nature of the cloning vehicle, a productive DNA insert should be expressed as a part of a larger fusion polypeptide. To verify this prediction, we cloned into pMR100 sonicated DNA (carefully size fractionated to avoid cloning small DNA fragments) derived from a cloned piece of *Drosophila melanogaster* DNA. Eleven red colonies were chosen for detailed analysis; the data are summarized in Table 2. Five of the 11 colonies contain proteins that are bigger (by approximately 20,000 daltons) than the initial cI–lacIZ fusion protein, consistent with a DNA insert size of approximately 500 bp (Table 2 and Fig. 3). These large proteins react with both anti-cI and anti-β-galactosidase antisera (Table 2 and Fig. 4), consistent with the hypothesis that they are proteins of the general structure cI–eukaryotic piece–lacIZ. These five strains contain additional smaller proteins, the largest of which is approximately 20,000 daltons less than the initial cI–lacZ fusion protein of pMR200. This protein reacts positively with anti-β-galactosidase antiserum but does not react with anti-cI antiserum. A protein with similar properties is also visible in the strain with pMR200. Two of the colonies (nos: 2 and 4) have properties suggesting that they are lacZ$^+$ revertants of pMR100 generated by a deletion of perhaps 1 bp during cloning. Four of the colonies (nos. 7, 9, 10, and 11) contain only the smaller proteins

Table 2. Insertion of cloned sonicated *Drosophila* DNA into pMR100

	Insert size	β-Gal activity	Protein size*	Protein blot	
				anti-cI	anti-β-Gal
Clone					
1	425	1,464	Small, big†	Big	Small, big
2	—	ND	pMR200	ND	ND
3	515	1,547	Small, big	Big	Small, big
4	—	ND	pMR200	ND	ND
5	425	1,569	Small, big	Big	Small, big
6	520	1,885	Small, big	Big	Small, big
7	460	660	Small	—	Small
8	440	795	Small, big	Big	Small, big
9	400	762	Small	—	Small
10	530	441	Small	—	Small
11	400	351	Small	—	Small
LG90	—	2	—	ND	ND
pMR100	—	30	—	—	—
pMR200	—	10,661	pMR200	Big	Small, big

A *Drosophila melanogaster* clone, pDm2837, was a generous gift of Welcome Bender. This eukaryotic insert, an 8.1-kilobase (kb) *Sal* I fragment in the *Sal* I site of pBR322, has been identified on the basis of genetic criteria to include the ry^+ gene, which codes for xanthine dehydrogenase (Welcome Bender and Arthur Chovnick, personal communication). The 8.1-kb *Sal* I insert was purified by agarose gel electrophoresis and electroelution and then sonicated. DNAs of 400–550 bp were selected for cloning, and 11 red colonies were chosen for subsequent analysis. No colonies or data were discarded from the analysis. β-Gal, β-galactosidase; —, no detectable insert or protein band; ND, not done.

* Protein size refers to the prominent β-galactosidase polypeptide(s) visible on $NaDodSO_4$ gels as assayed by Coomassie blue staining and blotting with anti-β-galactosidase antibody.

† "Big" refers to a β-galactosidase polypeptide 15,000–20,000 daltons larger than the cI–lacIZ fusion in pMR200. "Small" refers to a β-galactosidase polypeptide 15,000–20,000 daltons smaller than the cI–lacIZ fusion in pMR200; none of these "small" proteins reacts with anti-cI-antibody.

(no. 7 is shown in Figs. 3 and 4). The properties of these polypeptides are identical to those of the small protein of the other clones described above. The origin of this small protein is uncertain (see *Discussion*).

DISCUSSION

The experiments presented here validate our initial hypothesis: a plasmid vector having a frameshift near the amino terminus of a gene that codes for a polypeptide with $lacZ^+$ activity can be used to identify, clone, and express open reading frame segments of DNA. Although other methods are available to carry out these functions, they usually require detailed information, often at the DNA sequence level, about the gene or genes of interest. The value of the vector and approach described here is that *no* information need be known to clone and express fragments of open reading frame DNA from among a larger number of fragments.

The data presented have been chosen from a large number of similar experiments to illustrate some important features of the methodology. When shotgun-cloning random fragments of DNA, it is important to size the DNA carefully prior to ligation. Because the frequency of random open reading frames increases in inverse proportion to DNA fragment size, the selection of red colonies will also select for small DNA inserts, even if the average DNA size is quite large (data not shown). In the absence of any information about the DNA to be cloned in this way, we have chosen a DNA insert size of 400–500 bp. The probability that a random piece of DNA of length x is the correct size ($3n$

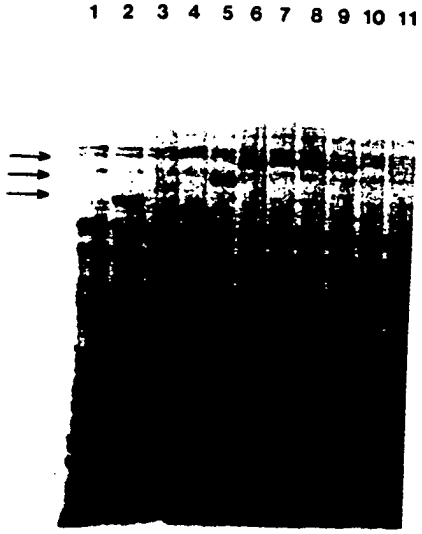


FIG. 3. Gel analysis of hybrid proteins. Proteins from various sources were analyzed by $NaDodSO_4$ gel electrophoresis and the gel was stained with Coomassie blue. Lanes: 1, HB101; 2, strain HB101 was grown overnight at 37°C in the presence of 1 mM isopropyl-β-D-thiogalactoside; 3, LG90; 4, pMR100; 5, pMR200; 6–11, clones 1, 3, and 5–8, respectively (see Table 2). Arrows: lower, wild-type lacZ (lane 2); middle, cI–lacIZ fusion protein (lane 5); upper, large insert-derived fusion proteins (lanes 6–9 and 11).

$- 1$) and entirely open reading frame $\approx (1/3) \times (0.953)^x \approx 5.5 \times 10^{-4}$ for 400-bp DNA and 1.1×10^{-4} for 500-bp DNA. The probability that a segment of DNA that is derived from a larger piece of open reading frame DNA is of the correct length and entirely open reading frame = the probability that it is the correct length $(1/3) \times$ the probability that it is in the correct frame $(1/3) \times$ the probability that it has cloned in the correct orientation $(1/2) = 5.5 \times 10^{-2}$. The difference between these numbers suggests that an open reading frame clone from a 400- to 500-bp DNA insert can be considered with some confidence to

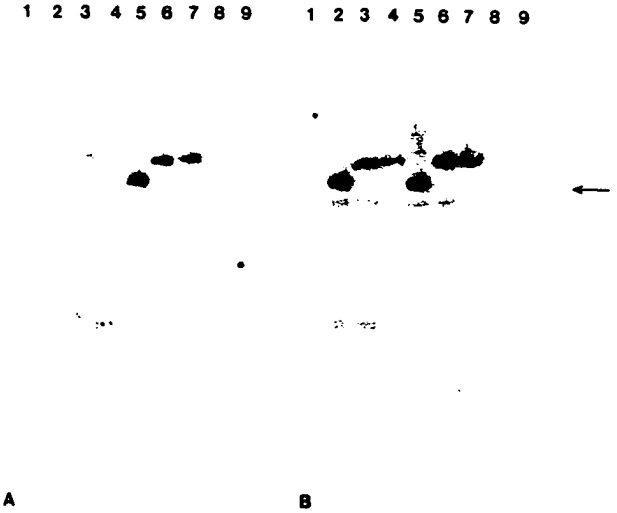


FIG. 4. Blot analysis of hybrid proteins. Proteins from various sources were analyzed by protein blotting and antibody staining with anti-cI antibody (*A*) and anti-β-galactosidase antibody (*B*). Lanes: 1, pMR100; 2, pMR200; 3–9, clones 1 and 3–8, respectively (see Table 2). The arrow indicates the small protein discussed in the text. The additional smaller protein bands present in all lanes (except lane 1) are probably fusion protein degradation products that lack discrete portions from the carboxyl terminus, since these bands are visible with anti-cI antibody.

be a bona fide open reading frame in the DNA from which it was derived. This level of confidence decreases considerably as a function of insert size. Such an argument suggests that genes with small exons and large introns are not generally approachable in this way although sonicated double-stranded cDNA could be cloned in a similar fashion.

The data in Table 1 argue that the strategy enriches significantly for open reading frame DNA segments. On the other hand, some false positives are certainly generated, as shown in Table 2 and Figs. 3 and 4. Some red colonies contain plasmid DNA with no detectable DNA insert and a lacZ polypeptide of a molecular weight identical to the initial cI–lacIZ fusion protein. As indicated above, these colonies are probably due to a small amount of exonuclease activity present during *Sma* I digestion or ligation. This source of red colonies is more problematic when the fraction of bona fide open reading frame colonies is low [e.g., when cloning genomic DNA from higher eukaryotes (unpublished experiments)]. There are also red colonies with proper DNA inserts but no detectable large fusion proteins. We have generated colonies of this phenotype from many sources of eukaryotic DNA (data not shown). They are always associated with the presence of a small lacZ polypeptide of molecular weight similar or identical to the lacIZ-encoded portion of the fusion protein. Because a similar lacZ polypeptide is visible in protein from pMR200 and all of these small polypeptides fail to react with anti-cI antibody, we favor the interpretation that they are due to proteolysis of the larger cI-positive fusion protein. Presumably, this occurs to a variable extent with different fusion proteins. Alternatively, these small proteins from the insert DNA-containing plasmids may be due to translational restarts near the insert–*lacIZ* junction. This matter might be clarified by determination of the DNA sequence of several inserts from plasmids of this type.

Despite these two types of fusion proteins, a substantial fraction of red colonies contain bona fide open reading frame inserts as determined by the nature of the fusion protein they generate (Fig. 4). We expect this general approach to prove useful for a number of purposes. In addition to those described here, many of the fusion proteins generated in this or in a similar way may be antigenic and immunogenic. Large numbers of lac$^+$ colonies could be selected on lactose minimal plates and screened wit immunological reagents. Thus, it should be feasible to clon specific genes on this basis. The methodology should also serv as a rapid and convenient way to proceed in the opposite di rection [i.e., from a gene (or a subgene region) to an antibod reagent directed against a portion of the proteins coded for b the starting DNA].

We are grateful to the individuals who supplied reagents. Also, w thank Drs. Donald Winkelmann and L. E. Cannon for advice abou protein blotting and M. A. Osley and L. M. Hereford for helpful dis cussions. This work was supported by National Institutes of Healt Grant HD-08887 (to M.R.). M.R.G. was supported by a predoctora Kessner Fellowship.

Note Added in Proof. The large proteins from clones 5 and 6 react pos itively with antixanthine dehydrogenase antiserum (a gift of Arthu Chovnick), showing that the eukaryotic protein pieces are antigenic.

1. Bassford, P., Beckwith, J., Berman, M., Brickman, E., Casada ban, M., Guarente, L., Saint-Girous, I., Sarthy, A., Schwartz M. & Silhavy, T. (1978) in *The Operon*, eds. Miller, J. H. & Rez nikoff, W. S. (Cold Spring Harbor Laboratory, Cold Spring Har bor, NY), pp. 245–262.
2. Guarente, L. & Ptashne, M. (1981) *Proc. Natl. Acad. Sci. USA* 78. 2199–2203.
3. Rose, M., Casadaban, M. J. & Botstein, D. (1981) *Proc. Natl Acad. Sci. USA* 78, 2460–2464.
4. Miller, J. H. (1972) *Experiments in Molecular Genetics* (Colc Spring Harbor Laboratory, Cold Spring Harbor, NY).
5. Clewell, D. B. & Helinski, D. R. (1969) *Proc. Natl. Acad. Sci USA* 62, 1159–1166.
6. Ish-Horowicz, D. & Burke, J. F. (1981) *Nucleic Acids Res.* 9 2989–2998.
7. Maniatis, T., Jeffrey, A. & Van de Sande, H. (1975) *Biochemistry* 14, 3787–3794.
8. Dagert, M. & Ehrlich, S. D. (1979) *Gene* 6, 23–28.
9. Laemmli, U. (1970) *Nature (London)* 227, 680–685.
10. Towbin, H., Staehelin, T. & Gordon, J. (1979) *Proc. Natl. Acad. Sci. USA* 76, 4350–4354.
11. Backman, K., Ptashne, M. & Gilbert, W. (1976) *Proc. Natl. Acad Sci. USA* 73, 4174–4178.
12. Guarente, L., Lauer, G., Roberts, T. M. & Ptashne, M. (1980) *Cell* 20, 543–553.
13. Sauer, R. T. (1978) *Nature (London)* 276, 301–302.

Exhibit 2

Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 3675–3679, June 1984
Biochemistry

Method to map antigenic determinants recognized by monoclonal antibodies: Localization of a determinant of virus neutralization on the feline leukemia virus envelope protein gp70

(recombinant DNA/λ phage library/screening/fusion protein/epitope)

J. H. NUNBERG, G. RODGERS, J. H. GILBERT, AND R. M. SNEAD

Department of Molecular and Biological Research, Cetus Corporation, Emeryville, CA 94608

Communicated by Hamilton O. Smith, March 8, 1984

ABSTRACT **A method is presented whereby antigenic determinants recognized by specific monoclonal antibodies can be mapped to specific sites on a protein sequence with high resolution. Short DNase I-generated DNA fragments encoding portions of the protein of interest are molecularly cloned into the *Eco*RI site of the β-galactosidase gene of phage λ Charon 16 so as to obtain expression of random protein fragments as fusion proteins. The monoclonal antibody is used to screen the phage library to isolate phage expressing the specific antigenic determinant. DNA of immunoreactive phage can be analyzed rapidly and subcloned to allow DNA sequence determination. The method is generally applicable and permits antigenic determinants of functionally interesting monoclonal antibodies to be mapped and related to specific protein sequences. We have used this procedure to determine the region of the feline leukemia virus envelope protein gp70 recognized by a virus-neutralizing monoclonal antibody, cl.25. Antibody binding was mapped to a 14-amino acid region in the amino-terminal half of gp70. This region may be directly involved in an essential function of the gp70 protein, perhaps in gp70-mediated host recognition functions. Synthetic peptides derived from this region may provide useful vaccine antigens for the prevention of feline leukemia virus-associated disease in cats.**

Monoclonal antibodies (mAbs) have been invaluable for the understanding of structural and functional features of viral proteins. Epitopes recognized by specific mAbs define structural and immunological domains of proteins and serve, operationally, as units of protein structure. For only a small number of viral proteins, however, has it been possible to relate operationally defined mAb epitopes to specific protein sequences. Current methods used to define sites of antibody binding often involve low-resolution mapping to specific chemical or proteolytic fragments of the protein (1–3). Sequence analysis of naturally arising viral variants that differ antigenically has been widely used to correlate protein sequence and immunologic differences (4, 5). This method has recently been extended to include the analysis of viral variants selected for resistance to specific virus-neutralizing mAbs; epitopes involved in virus neutralization have been mapped precisely by this procedure (6–8). The use of synthetic peptides to generate specific polyclonal antibodies has also permitted the identification of functionally interesting domains within proteins (9).

In this paper, we present a generally applicable high-resolution method to map specific antigenic determinants recognized by mAbs. Specific antibodies are used to screen recombinant phage libraries expressing, as fusion proteins, short random fragments of sequences encoding the relevant protein. The use of antibody probes to detect expression of a variety of specific antigens in recombinant libraries has been described by others (10–12). Recently, Young and Davis (13) have described a method for high-density screening of phage libraries containing genomic DNA from *Saccharomyces cerevisiae*. Polyclonal antibodies directed against two subunits of RNA polymerase II were used to identify phage expressing, as fusion proteins, antigens derived from the respective genes. Ruther *et al.* (14) have used polyclonal antibodies to isolate lysozyme-encoding sequences from a plasmid library containing small random fragments derived from molecularly cloned genomic sequences; the smallest such fragment identified encoded 10 amino acids of lysozyme. The method presented here uses high-density phage-screening procedures to identify, within proteins, small immunogenic regions recognized by specific mAbs.

We have used this method to localize a region of the feline leukemia virus (FeLV) envelope protein gp70 recognized by a specific virus-neutralizing mAb. The FeLVs comprise a group of horizontally transmitted retroviruses associated with a variety of naturally occurring malignant and degenerative diseases in domestic cats (15–17). Products of the viral envelope gene, specifically the major envelope glycoprotein gp70, are believed to be involved in recognition functions relating to virus host range (18, 19) and to serve as primary targets for antibody-mediated virus neutralization (20, 21). FeLV gp70 protein is found on the viral envelope in association with the hydrophobic anchor protein p15E; both proteins are generated during virus maturation by cleavage of the gp85 precursor protein found on the surface of infected cells.

In this paper, we report the mapping of the antigenic determinant recognized by a virus-neutralizing mAb to a region of 14 amino acids within FeLV gp70. The ability to localize, on the amino acid sequence level, epitopes recognized by functionally interesting mAbs should prove useful in correlating protein structure and function.

MATERIALS AND METHODS

Monoclonal Antibodies. mAbs directed against FeLV gp70 were provided by N. Pedersen (University of California, Davis). gp85 complex composed of cross-linked gp70 and p15E was prepared as described by Schneider *et al.* (22) from FL-74 FeLV and used to immunize mice. Hybridomas were prepared as described by Lutz *et al.* (23). FL-74 FeLV contains a mixture of all FeLV subgroups (A, B, and C) (18).

Virus neutralization was determined in a focus-induction assay using CCC clone 81 (S^+L^-) cells (24) or by virus-yield reduction using PG-4II(E) (S^+L^-) cells provided by D. K.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: FeLV, feline leukemia virus; GA-FeLV, Gardner–Arnstein isolate of FeLV; MuLV, murine leukemia virus; mAb, monoclonal antibody; bp, base pair(s); Ch.16, phage λ Charon 16; $\bar{M}_m$ and $\bar{M}_n$, mass and number, respectively, average molecular weight; X-Gal, 5-bromo-4-chloro-3-indolyl β-D-galactoside.

Haapala (Frederick Cancer Research Institute); virus was preincubated with antibody (30 min at 37°C) and allowed to infect cells as described by Fischinger *et al.* (24). Virus yield was determined by p27 ELISA (23). All sera and antibody preparations were heated to inactivate complement.

Construction of Phage Library. Molecularly cloned FeLV-B [Gardner–Arnstein (GA) isolate] provirus (λHF60) (25) was provided by A. Roach and N. Davidson (California Institute of Technology) in the form of the pKC7-derived plasmid pKHR-1. Short random segments of the GA-FeLV envelope gene were used to construct a library in λ phage Charon 16 (Ch.16) (26); fragments were derived from the FeLV envelope gene by digestion with DNase I in the presence of $MnCl_2$ (27). The *Xho* I/*Eco*RI DNA fragment of pKHR-1 containing the envelope gene was incubated with DNase I (Worthington) (100 μg of DNA/10 ng of DNase I per ml) at 25°C, and time points were analyzed by ethidium bromide staining of agarose gels to determine mass average molecular weight ($\bar{M}_m$). $\bar{M}_m$ values of 500–700 base pairs (bp) were obtained by 10–30 min digestion and correspond, in our experience, to number average molecular weight ($\bar{M}_n$) values of approximately 100 bp as determined by analysis of end-labeled fragments. Fragments were made blunt-ended by treatment with DNA polymerase I (Klenow fragment) in the presence of dNTPs and ligated to phosphorylated *Eco*RI linkers (5′ G-G-A-A-T-T-C-C 3′; Collaborative Research, Waltham, MA) as described by Maniatis *et al.* (28). The ligation mixture was digested with *Eco*RI and fractionated on cellulose (CF-11; Whatman) as described by Franklin (29) to remove linker monomer, residual triphosphate, and other contaminants. An estimate of actual $\bar{M}_n$ could be obtained at this point.

The *Eco*RI site in the β-galactosidase gene of Ch.16 was used to construct an expression library of random envelope gene fragments. The library was constructed (30) and packaged *in vitro* (31), as described by Maniatis *et al.* (28). Phage vector DNA was dephosphorylated using bacterial alkaline phosphatase prior to ligation to maximize insertion frequency. The resulting library was amplified prior to screening.

Immunoscreening. Phage were grown in the LacZ$^+$ *Escherichia coli* K-12 strain LE392 (28) or in the LacZ$^-$ strain MC1000 (32) on plates (26) containing the chromogenic indicator 5-bromo-4-chloro-3-indolyl β-D-galactoside (X-Gal). After overnight growth, a dry nitrocellulose filter (Schleicher & Schuell, BA85) was gently placed on the plate and marked for orientation. Phage were allowed to adsorb for 30–60 min and filters were screened much as described by Young and Davis (33). Filters were rinsed in 10 mM Tris·HCl, pH 7.4/150 mM NaCl/1 mM EDTA/0.1% $NaDodSO_4$ for 30 min and then incubated in 10 mM Tris·HCl, pH 7.4/150 mM NaCl/1 mM EDTA/0.2% Triton X-100/0.1% $NaDodSO_4$ (buffer B), containing 3% ovalbumin (Sigma) for 60 min to block nonspecific protein binding sites. These and all subsequent steps were carried out at room temperature. Incubations were carried out in polyethylene bags containing 3.5 ml of solution per filter. After incubation, the filters were allowed to react with virus-neutralizing mAb (50 μg of IgG/ml) in buffer B/1% ovalbumin for 3 hr and washed three times (5 min each) in the same buffer prior to incubation with rabbit antimouse IgG antibody (10 μg of IgG/ml) (Cappel Laboratories, Cochranville, PA) for 60 min. This second antibody had been adsorbed against *E. coli* proteins (13) to reduce nonspecific binding. Optimal antibody concentrations were determined to maximize the signal-to-noise ratio. After three more washes in the same buffer, bound antibody was detected by reaction with ^{125}I-labeled *Staphylococcus aureus* protein A (5×10^7 cpm/μg) for 60 min (or overnight at 4°C) in buffer B/1% ovalbumin (5×10^6 cpm per filter). Filters were finally washed four times (5 min each) in this buffer, dried, and autoradiographed. Reliable signals were obtained overnight. Immunoreactive phage were picked and purified by repeated screenings.

Nucleic Acid Analysis of Immunoreactive Phage. DNA from immunoreactive phage was purified as described by Maniatis *et al.* (28). The size and number of random envelope gene fragments inserted within reactive phage DNA was determined by digestion with *Eco*RI. *Eco*RI-digested DNA was end-labeled using DNA polymerase I (Klenow fragment) and analyzed by polyacrylamide gel electrophoresis to resolve inserted fragments.

To determine the DNA sequence of inserted random envelope fragments, an *Sst* I/*Pst* I DNA fragment from immunoreactive phage containing the *Eco*RI cloning site as well as flanking β-galactosidase sequences (see Fig. 3) was subcloned into plasmid pUC13 (34). This fragment [≈3.5 kilobase pairs plus insert(s)] was easily isolable by molecular cloning from *Pst* I- and *Sst* I-digested phage DNA by virtue of the unique *Sst* I site. Phage DNA digested with *Pst* I and *Sst* I was ligated into similarly digested pUC13, and "white" ampicillin-resistant transformants were selected for restriction endonuclease analysis of plasmid DNA. Digestion with *Eco*RI proved informative in this analysis. DNA sequence analysis was by the procedures of Maxam and Gilbert (35); a restriction map of *lac* operon sequences within Ch.16 was determined from the previously published DNA sequence of the *lacz* and *lacy* genes (36, 37). Only a portion of the *lacy* gene is contained within Ch.16.

Other Procedures. Molecular cloning techniques used were as described in Maniatis *et al.* (28). Nucleic acid enzymes were obtained from New England Biolabs or Bethesda Research Laboratories and used according to the manufacturers' instructions. Immunological blotting analysis of expressed proteins was carried out as described by Erlich *et al.* (38); proteins were resolved on $NaDodSO_4$/polyacrylamide gels and transferred electrophoretically to cyanogen bromide-activated paper. mAb cl.25 binding was detected by using ^{125}I-labeled protein A (*S. aureus*) after reaction with rabbit anti-mouse IgG antibody. All buffers used in immunologic probing contained 0.1% $NaDodSO_4$/0.2% Triton X-100 as described above. Direct ELISA assays for antigen were carried out as described by Voller *et al.* (39); bacterial proteins were solubilized by boiling in buffer containing 2% $NaDodSO_4$/50 mM dithiothreitol prior to dilution into plate-coating buffer. Synthetic oligopeptides were synthesized using a SAM II automated peptide synthesizer (Biosearch, San Rafael, CA).

RESULTS

Virus-Neutralizing mAb cl.25. Several mAbs directed against the envelope glycoprotein gp70 of FeLV were examined for the ability to neutralize FeLV in a focus-induction assay using CCC cl.81 (S$^+$L$^-$) cells. One antibody, cl.25 (IgG), was found to neutralize most FeLV isolates tested, of all FeLV subgroups. Viruses neutralized include CT600 FeLV-A (40), CF-927 FeLV-A (41), GA-FeLV-B (25), Snyder–Theilen FeLV-B (18), and CF-927 FeLV-C (41). CF-927 FeLV-B (41) was not neutralized. We were interested in determining the site on FeLV gp70 involved in cl.25-mediated virus neutralization.

Construction of a Random DNA Fragment Library Expressing FeLV gp70 Determinants. Our approach to the identification of the antigenic determinant of FeLV gp70 recognized by the mAb cl.25 involved the construction of a phage library containing random small fragments of the FeLV envelope gene so as to obtain expression of random envelope antigens as fusion products with β-galactosidase. Phage expressing the specific antigenic determinant recognized by cl.25 were detected by immunoscreening using the antibody.

Having demonstrated mAb cl.25-mediated neutralization

of the Gardner–Arnstein (GA) isolate of FeLV-B, we constructed a random DNA fragment library from envelope gene sequences of molecularly cloned GA-FeLV-B (25). A 3.1-kilobase-pair *Xho* I/*Eco*RI DNA fragment containing the entire 2.0 kilobase pairs of envelope-encoding sequences was isolated and digested with DNase I to a mass average molecular weight of 500–700 bp. In our experience, this corresponds to a $\bar{M}_n$ of ≈100 bp, allowing for expression of, on the average, 33 amino acid segments of the FeLV envelope protein.

Random DNA fragments were molecularly cloned into the *Eco*RI site encoding amino acids 1007-1008 (Glu-Phe) of the β-galactosidase gene in Ch.16. Approximately 30,000 independent plaques were derived; roughly 20% contained inserts as judged by *in situ* plaque hybridization (42) using the original *Xho* I *Eco*RI GA-FeLV DNA fragment as probe (data not shown). From this library, ≈2000 independent phage were amplified and used in immunoscreening procedures. Using the simplest of assumptions,* a library of this size would be expected to yield two independent isolates expressing any specific determinant.

Immunoscreening Using mAb cl.25. To assess our ability to detect expression in *E. coli* of a β-galactosidase fusion protein containing the antigenic determinant recognized by cl.25, a recombinant phage, Ch.16:GA, was constructed in which sequences encoding the entire GA-FeLV gp85 precursor were fused, in translational reading frame, to the *Eco*RI site of Ch.16. Using sequences derived from a synthetic polylinker (5′ G-A-A-T-T-C-C-C-G-G 3′), we adapted the *Bal* I site encoding the FeLV envelope signal peptide gp85 junction (43, 44) to create an *Eco*RI site and allow an in-frame fusion with the *Eco*RI site of β-galactosidase. Uncloned phage from the *in vitro* packaging reaction were screened for expression of the cl.25 determinant as described and the results are shown in Fig. 1*A*. Purified phage were shown, by restriction endonuclease analysis, to contain the expected FeLV sequences in the expected orientation and were also shown to produce a fusion protein detectible by mAb cl.25 by immunological blotting analysis (Fig. 2).

These experiments define conditions for successful screening and demonstrate the ability of the specific mAb to recognize the bacterially produced antigenic determinant. Antibodies directed against carbohydrate moieties of glycoproteins or complex conformational determinants of native proteins would be expected to fail to recognize the bacterial fusion protein.

Initial immunological screening of the amplified DNA fragment library was carried out at a plaque density of $\approx 10^3$ per 85-mm plate. Of 11 plaques chosen for examination, all but one gave consistently reliable signals on repeated screening and purification. The purification of one such phage (Ch.16:7) is shown in Fig. 1*B*. Purified phage were homogeneous as judged by immunoreactivity and X-Gal staining (faint blue on *E. coli* strain MC1000).

Analysis of Immunoreactive Phage DNA. DNA was isolated from immunoreactive phage and the structure was examined by restriction endonuclease analysis. Phage DNAs were digested with *Eco*RI to liberate the insert fragment(s), and fragments were labeled at the *Eco*RI ends by using DNA polymerase I (Klenow fragment). Polyacrylamide gel analysis showed all immunoreactive isolates to be identical: all contained one 51-bp *Eco*RI insert fragment. This result is consistent with the limited complexity of the amplified library.

Immunologic analysis of proteins expressed by one such immunoreactive phage (Ch.16:9-3) showed the presence of a fusion protein recognized by both mAb cl.25 and rabbit anti-β-galactosidase antibody (Fig. 2). These results confirm the immunologic specificity of the screening procedure. The fusion protein migrated slightly slower on $NaDodSO_4$/polyacrylamide gel electrophoresis than did the wild-type Ch.16 β-galactosidase.

To facilitate DNA sequence analysis of the 50-bp fragment involved in immunoreactivity, a region of phage DNA containing the *Eco*RI insertion site from Ch.16:9-3 was subcloned into plasmid pUC13. The DNA sequence of the envelope gene fragment insert was determined by the chemical sequencing procedures of Maxam and Gilbert; sequence was read from the *Nde* I site 50 bp from the *Eco*RI cloning site. For the analysis of more complex libraries, an oligonucleo-

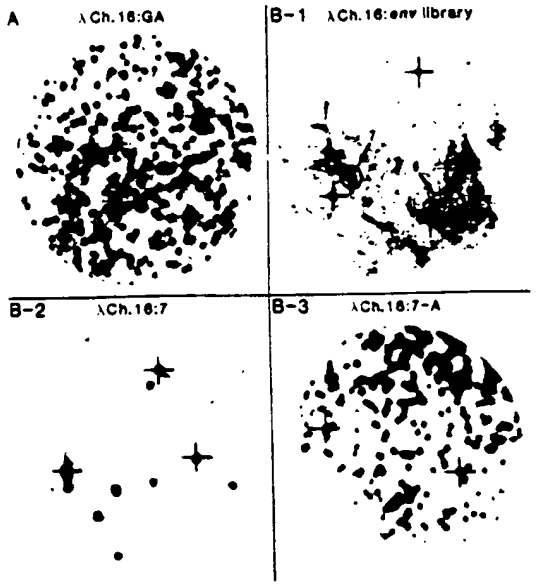


FIG. 1. Immunological screening of phage libraries for detection of expression of the antigenic determinant recognized by mAb cl.25. (*A*) Phage generated in the construction of Ch.16:GA were screened using the mAb. Ch.16:GA contains sequences of the mature FeLV gp85 fused, in frame, to β-galactosidase. (*B*) Purification of an immunoreactive phage (Ch.16:7) from a library containing random FeLV envelope gene fragments. After the initial identification (B-1), phage were purified by repeated immunological screening (B-2) to final homogeneity (B-3). Purified phage were homogeneous by immunological screening and by X-Gal staining. Filter orientation markers are indicated by crosses.

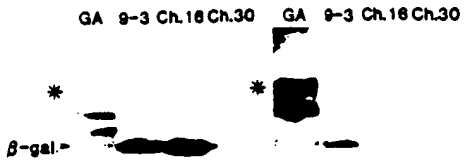
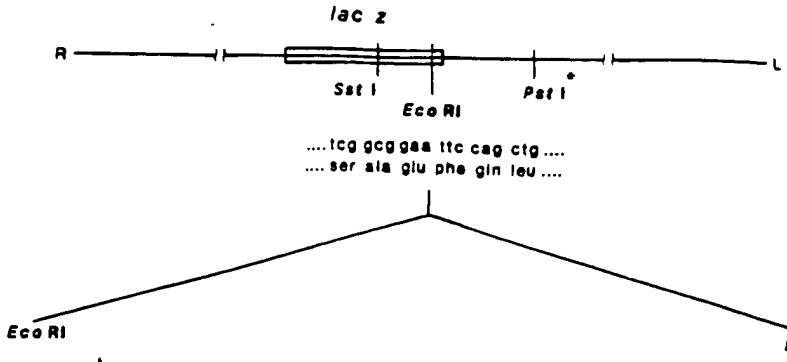


FIG. 2. Immunological blotting analysis of protein from immunoreactive phage. Protein was isolated from *E. coli* MC1000 cells containing Ch.16:GA (GA), Ch.16:9-3 (9-3), Ch.16, or Charon 30 and analyzed using, as probe, either rabbit anti-β-galactosidase IgG (*Left*) or mAb cl.25 (*Right*). Proteolytic degradation products of the intact Ch.16:GA fusion protein (*) are visible. Charon 30 does not contain the β-galactosidase gene and is included to demonstrate antibody specificity. β-gal, Native β-galactosidase.

*Estimation of expected frequency is derived as follows: 0.20 (insertion frequency) × [100 bp (M_n of insert)/3100 bp (complexity of DNA used to construct library)] × 0.5 (orientation) × 0.33 (reading frame).



λ Charon 16: 9-3
.... tcg gcg gaa ttc cTC ATG GGA CCA AAT CTA GTC CTG CCT GAT CAA AAA CCC CCA TCg gaa ttc cag ctg
.... ser ala glu phe leu MET GLY PRO ASN LEU VAL LEU PRO ASP GLN LYS PRO PRO SER glu phe gln leu....

GA-FeLV-B gp70
213 226
.... THR PRO PRO GLN ALA MET GLY PRO ASN LEU VAL LEU PRO ASP GLN LYS PRO PRO SER ARG GLN SER GLN....

FIG. 3. DNA sequence analysis of immunoreactive phage Ch.16:9-3. Ch.16 DNA is diagrammed to indicate the position of the β-galactosidase gene (*lacz*) and the *Eco*RI cloning site. Left and right "arms" of the phage DNA are indicated. The nucleic acid and deduced amino acid sequences surrounding the *Eco*RI site are indicated below. The nucleic acid and deduced amino acid sequences present in Ch.16:9-3 are indicated in relation to Ch.16 sequences. The 43-bp insert derived from the envelope gene of GA-FeLV-B is indicated in capital letters, as are the deduced envelope gene-encoded amino acids. (Sequences derived from the synthetic *Eco*RI linkers flank the insert.) The comparable region of GA-FeLV gp70 protein is indicated below. *, The *Pst* I site referred to in the text (other *Pst* I sites exist within the phage DNA); ↑, this thymidine represents a mutation from the parental GA-FeLV sequence, which contains a cytosine at this position (43, 44) (this mutation apparently arose during the molecular cloning of the random fragment library).

tide primer complementary to sequences adjacent to the *Eco*RI cloning site could be used to facilitate sequence determination from a large number of distinct pUC13 [or M13mp11 (45)] subclones using primer-dependent dideoxynucleotide sequencing methods (46, 47).

DNA sequence analysis of pUC13:9-3 demonstrated the presence of a 51-bp insert containing 43 bp derived from the envelope gene of Ga-FeLV (Fig. 3). This isolated envelope gene fragment represents nucleotides 635–678 of sequences encoding gp70 and is present so as to give rise to an in-frame fusion protein with flanking β-galactosidase sequences. The amino acid sequence of the encoded gp70 fragment (residues 213–226 of gp70) is also shown in Fig. 3. Thus, the antigenic determinant recognized by virus-neutralizing mAb cl.25 is contained within this 14-amino acid region.

To further demonstrate the presence of the mAb cl.25 determinant within this fragment, the 51-bp *Eco*RI fragment of pUC13:9-3 was recloned into the *Eco*RI site of an expression system plasmid containing the *trpLE'* gene (48). The *Eco*RI sites of β-galactosidase and *trpLE'* exist in the identical reading frame (Glu-Phe), ensuring in-frame translation of the properly oriented 51-bp fragment with a *trpLE'* fusion protein. Transformants were screened using mAb cl.25 in an ELISA, and results were confirmed by immunological blotting analysis (data not shown); *trpLE'* fusion proteins containing this region of FeLV gp70 are also recognized by the mAb. The most direct indication that the antigenic determinant recognized by mAb cl.25 resides strictly within this 14-amino acid fragment was obtained through the use of a synthetic peptide containing this sequence. The peptide, Met-Gly-Pro-Asn-Leu-Val-Leu-Pro-Asp-Gln-Lys-Pro-Pro-Ser, was synthesized and shown to block the binding of mAb cl.25 to antigen in a gp70 ELISA and, in addition, to prevent mAb cl.25-mediated virus neutralization (data not shown).

DISCUSSION

The procedure described here appears to be generally applicable for the localization of antigenic determinants recognized by mAbs. The ability to isolate mAbs against specific proteins has been invaluable in defining and dissecting antigenic and functional epitopes of proteins. Current methods to localize antigenic determinants recognized by mAbs to specific protein sequences, however, are complicated and not of general utility. van der Werf and co-workers (49, 50) have recently described a procedure in which molecularly cloned sequences encoding a specific protein are expressed as a fusion protein in bacteria; internal deletions are then generated *in vitro* and correlated with loss of specific immunoreactivity. The method described here uses phage screening procedures to directly identify specific sequences encoding antigenic determinants of specific mAbs. Furthermore, this method requires no knowledge of the molecular target of the antibody; thus interesting proteins and antigenic determinants can be isolated from phage libraries derived from genomic DNA of more complex organisms from which important antigens have not yet been defined.

This procedure has been used to determine the region of FeLV gp70 recognized by the FeLV-neutralizing mAb cl.25. This region, comprising amino acids 213–226 of gp70, is identically conserved between the closely related Snyder-Theilen-FeLV-B and GA-FeLV-B viruses (44) and is quite closely conserved between feline and murine leukemia viruses (MuLVs) [e.g., 65% (9/14) homology to Moloney-MCF-MuLV gp70 (51)]. Isolated differences within this region exist among MuLVs (51–55). Virus host-range recognition functions are believed to reside within the amino-terminal half of gp70, and it is interesting to speculate that this region, identified by reaction with a neutralizing mAb, may be directly involved in this process. Hydropathic analysis, using the algorithm of Hopp and Woods (56), shows that this region lies within a local peak of hydrophilicity in gp70 (44) and thus, presumably, resides on the outside surface of the protein. It is interesting to note that this relatively well-conserved region overlaps, in part, the amino terminus of a highly variable type-specific region within MuLV gp70 proteins (52, 53, 55). It has been proposed that this latter region imparts type specificity on the MuLV gp70. These two regions may represent one functional protein domain involved in host recognition; alternatively, mAb cl.25 may act sterically to block access to an adjacent type-specific region.

The ability to define, on a viral protein, a site involved in antibody-mediated virus neutralization may serve to direct

the choice of synthetic peptides to be tested as vaccine antigens. Synthetic peptides to such a region of foot and mouth disease virus VP1 have been tested in the development of a synthetic peptide vaccine against the disease (9). Peptides derived from the region of FeLV gp70 identified here may provide useful vaccine antigens for the prevention of FeLV-associated disease in cats.

Note Added in Proof. Three independent FeLV-neutralizing mAbs derived by Osterhaus and Weijer were also shown to recognize the synthetic peptide. These mAbs had previously been shown to compete with mAb cl.25 for gp70 binding [A. Osterhaus (National Institute of Public Health, The Netherlands) and K. Weijer (The Netherlands Cancer Institute), personal communication].

We are indebted to Niels Pedersen (University of California, Davis, Veterinary Medicine) for providing us with monoclonal antibodies, viruses, and other reagents and to Frank McCormick for discussions from which this approach to mapping antigenic determinants was derived. We thank our colleagues at Cetus who were involved in aspects of this work: Glenn Horn for molecular cloning into Charon 16, Danute Nitecki for oligopeptide synthesis, Mark Williams for DNA sequence analysis, David Gelfand for critical review of the manuscript, and Tom White for encouragement and support. Molecularly cloned GA-FeLV and virus were provided by Norman Davidson (California Institute of Technology). This work was partly supported by funds from the TechAmerica Group, Inc.

1. Dietzschold, B., Wiktor, T. J., MacFarlan, R. & Varrichio, A. (1982) *J. Virol.* **44**, 595–602.
2. Bachrach, H. L., Morgan, D. O., McKercher, P. D., Moore, D. M. & Robertson, B. H. (1982) *Vet. Microbiol.* **1**, 85–96.
3. Jackson, D. C., Dopheide, T. A., Russell, R. J., White, D. O. & Ward, C. W. (1979) *Virology* **93**, 458–465.
4. Cheung, A., DeLamarter, J., Weiss, S. & Kupper, H. (1983) *J. Virol.* **48**, 451–459.
5. Webster, R. G., Laver, W. G., Air, G. M. & Schild, G. C. (1982) *Nature (London)* **296**, 115–121.
6. Caton, A. J., Brownlee, G. G., Yewdell, J. W. & Gerhard, W. (1982) *Cell* **31**, 417–427.
7. Minor, P. D., Schild, G. C., Bootman, J., Evans, D. M. A., Ferguson, M., Reeve, P., Spitz, M., Stanway, G., Cann, A. J., Hauptmann, R., Clarke, L. D., Mountford, R. C. & Almond, J. W. (1983) *Nature (London)* **301**, 674–679.
8. Evans, D. M. A., Minor, P. D., Schild, G. S. & Almond, J. W. (1983) *Nature (London)* **304**, 459–462.
9. Bittle, J. L., Houghten, R. A., Alexander, H., Shinnick, T. M., Sutcliffe, J. G. & Lerner, R. A. (1982) *Nature (London)* **298**, 30–33.
10. Erlich, H. A., Cohen, S. N. & McDevitt, H. O. (1978) *Cell* **13**, 681–689.
11. Broome, S. & Gilbert, W. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 2746–2749.
12. Hitzeman, R. A., Clarke, L. & Carbon, J. (1980) *J. Biol. Chem.* **255**, 12073–12080.
13. Young, R. A. & Davis, R. W. (1983) *Science* **222**, 778–782.
14. Ruther, U., Koenen, M., Sippel, A. E. & Muller-Hill, B. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 6852–6855.
15. Hardy, W. D., Jr., Old, L. J., Hess, P. W., Essex, M. & Cotter, S. (1973) *Nature (London)* **244**, 266–269.
16. Jarrett, O., Hardy, W. D., Jr., Golder, M. C. & Hay, D. (1978) *Int. J. Cancer* **21**, 334–337.
17. Pedersen, N. C., Theilen, G., Keane, M. A., Fairbanks, L., Mason, T., Orser, B., Chen, C.-H. & Allison, C. (1977) *Am. J. Vet. Res.* **38**, 1523–1531.
18. Sarma, P. S. & Log, T. (1973) *Virology* **54**, 160–169.
19. Sarma, P. S., Log, T., Jain, D., Hill, P. R. & Huebner, R. J. (1975) *Virology* **64**, 438–446.
20. Russell, P. H. & Jarrett, O. (1978) *Int. J. Cancer* **21**, 768–778.
21. De Noronha, F., Schafer, W., Essex, M. & Bolognesi, D. P. (1978) *Virology* **85**, 617–621.
22. Schneider, J., Falk, H. & Hunsmann, G. (1980) *J. Virol.* **33**, 597–605.
23. Lutz, H., Pedersen, N. C., Durbin, R. & Theilen, G. H. (1983) *J. Immunol. Methods* **56**, 209–220.
24. Fischinger, P. J., Blevins, C. S. & Nomura, S. (1974) *J. Virol.* **14**, 177–179.
25. Mullins, J. I., Casey, J. W., Nicolson, M. O., Bauman Burck, K. & Davidson, N. (1981) *J. Virol.* **38**, 688–703.
26. Blattner, F. R., Williams, B. G., Blechl, A. E., Denniston-Thompson, K., Faber, H. E., Furlong, L.-A., Grunwald, D. J., Kiefer, D. O., Moore, D. D., Schumm, J. W., Sheldon, E. L. & Smithies, O. (1977) *Science* **196**, 161–169.
27. Crouse, G. F., Simonsen, C. C., McEwan, R. N. & Schimke, R. T. (1982) *J. Biol. Chem.* **257**, 7887–7897.
28. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
29. Franklin, R. M. (1966) *Proc. Natl. Acad. Sci. USA* **55**, 1504–1511.
30. Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G. K. & Efstratiadis, A. (1978) *Cell* **15**, 687–701.
31. Sternberg, N., Tiemeier, D. & Enquist, L. (1977) *Gene* **1**, 255–280.
32. Casadaban, M. J. & Cohen, S. N. (1980) *J. Mol. Biol.* **138**, 179–207.
33. Young, R. A. & Davis, R. W. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1194–1198.
34. Vieira, J. & Messing, J. (1982) *Gene* **19**, 259–268.
35. Maxam, A. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
36. Kalnins, A., Otto, K., Ruther, U. & Muller-Hill, B. (1983) *EMBO J.* **2**, 593–597.
37. Buchel, D. M., Gronenborn, B. & Muller-Hill, B. (1980) *Nature (London)* **283**, 541–545.
38. Erlich, H. A., Rodgers, G., Vaillancourt, P., Araujo, F. G. & Remington, J. S. (1983) *Infect. Immun.* **41**, 683–690.
39. Voller, A., Bidwell, D. E. & Bartlet, A. (1979) *The Enzyme Linked Immunosorbent Assay (ELISA)* (Dynatech Laboratories, Alexandria, VA).
40. Rasheed, S. & Gardner, M. B. (1981) *J. Natl. Cancer Inst.* **67**, 929–933.
41. Rasheed, S. & Gardner, M. B. (1980) in *Feline Leukemia Virus*, eds. Hardy, W. D., Jr., Essex, M. & McClelland, A. J. (Elsevier/North-Holland, New York), pp. 393–400.
42. Benton, W. D. & Davis, R. W. (1977) *Science* **196**, 180–182.
43. Elder, J. H. & Mullins, J. I. (1983) *J. Virol.* **46**, 871–880.
44. Nunberg, J. H., Williams, M. E. & Innis, M. A. (1984) *J. Virol.* **49**, 629–632.
45. Messing, J. & Vieira, J. (1982) *Gene* **19**, 269–276.
46. Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. & Roe, B. A. (1980) *J. Mol. Biol.* **143**, 161–178.
47. Heidecker, G., Messing, J. & Gronenborn, B. (1980) *Gene* **10**, 69–73.
48. Kleid, D. G., Yansura, D., Small, B., Dowbenko, D., Moore, D. M., Grubman, M. J., McKercher, P. D., Morgan, D. O., Robertson, B. H. & Bachrach, H. L. (1981) *Science* **214**, 1125–1129.
49. van der Werf, S., Wychowski, C., Bruneau, P., Blondel, B., Crainic, R., Horodniceanu, F. & Girard, M. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 5080–5084.
50. Wychowski, C., van der Werf, S., Siffert, O., Crainic, R., Bruneau, P. & Girard, M. (1983) *EMBO J.* **2**, 2019–2024.
51. Bosselman, R. A., Van Straaten, F., Van Beveren, C., Verma, I. M. & Vogt, M. (1982) *J. Virol.* **44**, 19–31.
52. Shinnick, T. M., Lerner, R. A. & Sutcliffe, J. G. (1981) *Nature (London)* **293**, 543–598.
53. Koch, W., Hunsman, G. & Friedrich, R. (1983) *J. Virol.* **45**, 1–9.
54. Lenz, J., Crowther, R., Straceski, A. & Haseltine, W. (1982) *J. Virol.* **42**, 519–529.
55. Chen, R. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 5788–5792.
56. Hopp, T. P. & Woods, K. R. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 3824–3828.

Exhibit 3

Molecular Cloning

A LABORATORY MANUAL

T. Maniatis Harvard University
E. F. Fritsch Michigan State University
J. Sambrook Cold Spring Harbor Laboratory

Cold Spring Harbor Laboratory
1982

HYBRIDIZATION OF DNA OR RNA IMMOBILIZED ON FILTERS TO RADIOACTIVE PROBES

There are many methods available to hybridize radioactive probes in solution to DNA or RNA immobilized on nitrocellulose filters. These methods differ in the following aspects:

the solvent and temperature used (68°C in aqueous solution or 42°C in 50% formamide);
the volume of solvent and the length of hybridization (large volumes for periods as long as 3 days or minimal volumes for times as short as 4 hours);
the degree and method of agitation (continuous shaking or stationary);
the concentration of the labeled probe and its specific activity;
the use of compounds, such as dextran sulfate, that increase the rate of reassociation of nucleic acids;
the stringency of washing following the hybridization.

Although the choice depends to a large extent on personal preference, we would like to offer the following guidelines.

1. Hybridization reactions in 50% formamide at 42°C are easier to set up, present less of an evaporation problem, and are less harsh on the filters than is hybridization at 68°C in an aqueous solution. The kinetics of the hybridization reaction in 80% formamide are approximately three to four times slower than in an aqueous solution (Casey and Davidson 1977). Assuming a linear relationship between the rate of hybridization and formamide concentration, the rate in 50% formamide should be two times slower than in an aqueous solution.

2. The smaller the volume of hybridization solvent, the better. The kinetics of nucleic acid reassociation are faster, and the amount of probe needed may be reduced so that the DNA on the filter acts as the driver for the reaction. All these are important parameters when detecting clones of low-abundance mRNAs. However, it is essential that sufficient liquid be present for the filters to remain at all times covered by a film of the hybridization solution.

3. Continual movement of the probe solution across the filter is unnecessary, even for a reaction driven by DNA immobilized on the filter. However, if a large number of filters are hybridized simultaneously, agitation is advisable in order to prevent the filters from adhering to each other.

4. The kinetics of the hybridization reaction are difficult to predict from theoretical considerations, partly because the exact concentration of the immobilized nucleic acid and its availability for hybridization are unknown.

When using probes made by nick translation of double-stranded DNA, the following rule of thumb is useful: Allow the hybridization to proceed for a time sufficient to enable the probe in solution to achieve 1-3× $C_0t_{1/2}$. In 10 ml of hybridization solution, 1 μg of a probe of 5-kb complexity will reach $C_0t_{1/2}$ in 2 hours. To determine the time of half-renaturation for any other probe, simply enter the appropriate values into the following equation:

$$\frac{1}{X} \times \frac{Y}{5} \times \frac{Z}{10} \times 2 = \text{number of hours to achieve } C_0t_{1/2}$$

where,

X = the weight of probe added (in μg)
Y = its complexity (for most probes, complexity is proportional to the length of the probe in kb)
Z = the volume of the reaction (in ml)

After hybridization for 3× $C_0t_{1/2}$ has been reached, the amount of the probe available for additional hybridization to the filter is negligible. For single-stranded cDNA probes, the hybridization time may be shortened since the lack of a competing DNA strand in solution favors hybridization to DNA bound to the filter.

5. In the presence of dextran sulfate, the rate of association of nucleic acids is accelerated because the nucleic acids are excluded from the volume of the solution occupied by the polymer. Their effective concentration is therefore increased. The rate of association reportedly increases 10-fold in the presence of 10% dextran sulfate (Wahl et al. 1979).
   Although dextran sulfate is useful in circumstances where the rate of hybridization is the limiting factor in detecting sequences of interest, it is unnecessary for most purposes. It is also difficult to handle because of its viscosity and sometimes can lead to high backgrounds.

6. In general, the washing conditions should be as stringent as possible; i.e., a combination of temperature and salt concentration should be chosen that is slightly (5°C) below the T_m of the hybrid under study. The temperature and salt conditions can often be determined empirically in preliminary experiments where Southern blots (see pages 382ff) of genomic DNA are hybridized to the probe of interest and then washed under conditions of different stringency.

TABLE 11.1 HYBRIDIZATION CONDITIONS FOR SOUTHERN FILTERS

DNA on filter	Sp. act. of probe DNA (cpm/μg)	Amount of probe added	Time of hybridization (hr)
Fragments of cloned DNA (~ 100 ng/fragment)	10^7	10^5-10^6 cpm (0.01-0.1 μg)	3-4
Total eukaryotic DNA (10 μg)	10^8	1×10^7 cpm -5×10^7 (0.1-0.5 μg)	12-16

8. Incubate the bag submerged in a water bath at 68°C for the required hybridization period.

9. Remove the bag from the water bath and quickly cut along the length of three sides. Using gloves, remove the filter and immediately submerge it in a tray containing a solution of 2× SSC and 0.5% SDS at room temperature.

   Note. Do not allow the filter to dry out at any stage during the washing procedure.

10. After 5 minutes, transfer the filter to a fresh tray containing a solution of 2× SSC and 0.1% SDS and incubate for 15 minutes at room temperature with occasional gentle agitation.

11. Transfer the filter to a flat-bottomed plastic box containing a solution of 0.1× SSC and 0.5% SDS. Incubate at 68°C for 2 hours with gentle agitation. Change the buffer and continue incubating for a further 30 minutes.

    Note. If the homology between the probe and the DNA bound to the filter is inexact, the washing should be carried out under less stringent conditions. In general, washing should be carried out at $T_m - 12$°C.
    The following relationships are useful:

    a. $T_m = 69.3 + 0.41 \cdot (G + C)\% - 650/L$.
       L equals the average length of the probe in nucleotides (Marmur and Doty 1962; Wetmer and Davidson 1968).

    b. The T_m of a duplex DNA decreases by 1°C with every increase of 1% in the number of mismatched base pairs (Bonner et al. 1973).

c. $(T_m)\mu_2 - (T_m)\mu_1 = 18.5 \log_{10} \frac{\mu_2}{\mu_1}$

where μ_1 and μ_2 are the ionic strengths of two solutions (Dove and Davidson 1962).

12. Dry the filter at room temperature on a sheet of Whatman 3MM paper.

13. Wrap the filter in Saran Wrap and apply to X-ray film to obtain an autoradiographic image (see page 470).

Notes

Hybridization may also be carried out in:

a. flat-bottomed plastic boxes.

b. buffers containing formamide. Each increase of 1% in the formamide concentration lowers the T_m of a DNA duplex by 0.7°C (McConaughy et al. 1969; Casey and Davidson 1977).

Exhibit 4

Methods in Enzymology

Volume 100

Recombinant DNA

Part B

EDITED BY

Ray Wu
SECTION OF BIOCHEMISTRY
MOLECULAR AND CELL BIOLOGY
CORNELL UNIVERSITY
ITHACA, NEW YORK

Lawrence Grossman
DEPARTMENT OF BIOCHEMISTRY
THE JOHNS HOPKINS UNIVERSITY
SCHOOL OF HYGIENE AND PUBLIC HEALTH
BALTIMORE, MARYLAND

Kivie Moldave
DEPARTMENT OF BIOLOGICAL CHEMISTRY
COLLEGE OF MEDICINE
UNIVERSITY OF CALIFORNIA
IRVINE, CALIFORNIA

1983

ACADEMIC PRESS

A Subsidiary of Harcourt Brace Jovanovich, Publishers

New York London
Paris San Diego San Francisco São Paulo Sydney Tokyo Toronto

using the dot blot assay described here. This assay has been successfully used in several laboratories.[16,17]

Acknowledgment

The author thanks R. Reeder for helpful suggestions on this manuscript. This work was supported by a grant from the National Institutes of Health to Harold Weintraub. P. S. T. is a Postdoctoral Fellow of The National Institutes of Health.

[16] M. Groudine, R. Eisenman, and H. Weintraub, *Nature (London)* **292,** 311 (1981).
[17] S. F. Wolf and B. R. Migeon, *Nature (London)* **295,** 667 (1982).

[19] Isolation of Multigene Families and Determination of Homologies by Filter Hybridization Methods

By GERALD A. BELTZ, KENNETH A. JACOBS, THOMAS H. EICKBUSH, PETER T. CHERBAS, and FOTIS C. KAFATOS

A high proportion of genes in eukaryotes are now known to be members of multigene families. Although initial indications were provided by protein sequencing, the ubiquity of multigene families has been revealed most convincingly by recombinant DNA methods. A list of some of the best known multigene families, by no means inclusive, would include the genes for histones, globins, immunoglobulins, histocompatibility antigens, actins, tubulins, seed storage proteins, chorion proteins, keratins, collagens, cuticle proteins, yolk proteins, heat shock proteins, and salivary glue proteins. Even genes for proteins that appear homogeneous may belong to multigene families; e.g., the chicken ovalbumin gene is now known to be expressed coordinately with two homologous genes, *X* and *Y*. In the field of hormone research, analysis of gene families related to well known effectors such as insulin or growth hormone is one of the most exciting areas of current investigation. Many members of multigene families show stage- or tissue-specific expression: a familiar example is the family of globin genes. Thus, the study of multigene families is important for understanding the mechanisms of physiological regulation, differential gene expression, and molecular evolution in eukaryotes.

In the study of multigene families by recombinant DNA methods, usually the first step is to isolate clones (cDNA or genomic) that represent all members of the family. The next step is to discriminate among clones that correspond to different genes. The third step is to relate each gene to

ISBN 0-12-182000-9

a particular type of protein product or to a specific RNA transcript. Ultimately, structural characterization of the genes, including sequence analysis, is undertaken. Because of their sequence homologies, members of a multigene family often cross-hybridize. In the first three steps of such studies, careful control of the hybridization conditions is necessary—both to isolate the family members as cross-hybridizing species in the first step, and to permit their discrimination and characterization in the second and third steps. A number of filter hybridization methods are used in these studies: phage plaque[1] and bacterial colony hybridization,[2] dot blot hybridization,[3] Southern hybridization,[4] Northern hybridization,[5] and hybrid-selected translation.[6] Their specific features are discussed elsewhere in this volume. Here, we wish to emphasize their relatedness and to point out how deliberate control of the hybridization stringency maximizes their utility. We shall give examples from our studies of the chorion gene families in silkmoths.

The silkmoth chorion system has been reviewed.[7] More than a hundred genes belonging to several multigene families encode the structural proteins of the chorion. These families may themselves be related, constituting a superfamily,[8] but the interfamily homologies are low and can be neglected in this discussion. Within each family, both closely and distantly related genes are found: observed degrees of mismatching range from less than 1% to as much as 50%.[9] Based on this variation, members of each family have been classified into distinct *types* (distantly related genes, observed mismatching 10 to 50%) and into *copies* of the same gene type (more closely related genes, <1 to 5% mismatching). In our experience, even gene copies differ by one or more substitutions that lead to amino acid replacements, and thus have already taken the first step toward evolving into distinct genes, by the criterion of differences in the encoded proteins. The sequence variations are not uniformly distributed within the chorion genes: in each family the sequence corresponding to a central protein domain is substantially conserved, whereas flanking se-

[1] W. D. Benton and R. W. Davis, *Science* **196,** 180 (1977).
[2] M. Grunstein and D. S. Hogness, *Proc. Natl. Acad. Sci. U.S.A.* **72,** 3961 (1975).
[3] F. C. Kafatos, C. W. Jones, and A. Efstratiadis *Nucleic Acids Res.* **7,** 1541 (1979).
[4] E. M. Southern, *J. Mol. Biol.* **98,** 503 (1975).
[5] J. C. Alwine, D. J. Kemp, and G. R. Stark, *Proc. Natl. Acad. Sci. U.S.A.* **74,** 5350 (1977).
[6] R. P. Ricciardi, J. S. Miller, and B. E. Roberts, *Proc. Natl. Acad. Sci. U.S.A.* **76,** 4927 (1979).
[7] F. C. Kafatos, *Am. Zool.* **21,** 707 (1981).
[8] J. C. Regier, F. C. Kafatos, and S. J. Hamodrakas, *Proc. Natl. Acad. Sci. U.S.A.* (in press).
[9] C. W. Jones and F. C. Kafatos, *J. Mol. Evol.* **19,** 87 (1982).

quences that encode protein "arms" (NH_2-terminal and COOH-terminal domains) are much more variable.[10]

Parameters Affecting Nucleic Acid Hybridizations

A Qualitative Summary

By definition, the course of any hybridization reaction is determined by the concentrations of the reacting species and the second-order rate constant k. The stability of the resulting duplex is measured by the melting temperature, T_m. For reassociation reactions in solution involving perfectly matched complementary strands, the effects of various reaction conditions have been investigated in detail, and thus T_m and k values can be calculated with some precision. Neither calculation can be extended with confidence to cases involving imperfect hybrids or to solid-phase reactions. Nevertheless, the effects of reaction conditions can be expected to be qualitatively similar, and thus they bear review here.

In general, T_m depends on the composition of the duplex (T_m increases with G + C content), on the duplex length (T_m increases with length), on the ionic strength (T_m increases with salt concentration up to a plateau), and on the concentration of any organic denaturants present (e.g., formamide, urea). T_m is higher for RNA : DNA hybrids than for the corresponding DNA : DNA duplexes, and the differential effect of increasing (G + C) is greater for the hybrids.

Parameters such as (G + C), duplex length, ionic strength, and concentration of any denaturants generally affect k in the same direction as they affect T_m. Therefore, an additional parameter, which is of paramount importance for k, is the temperature of the reaction, T_i. In general, k increases with T_i, reaches a broad maximum at 25 to 20° below T_m, and decreases thereafter, becoming severely depressed at $T_i \geq T_m - 5°$. An example of such temperature dependence is shown in Fig. 1, for the reassociation of T4 DNA in solution.[11]

The qualitatively similar effects of most reaction conditions on k and T_m have led to the practice of summarizing the reaction conditions in terms of the *criterion,* which is numerically equal to T_m minus T_i. When the criterion is large, the reaction is described as permissive; when it is small, the reaction is stringent. While the *absolute* values of k vary drastically as a function of individual parameters (e.g., salt), the *relative* values as a function of criterion are thought to be described in general by the

[10] S. J. Hamodrakas, C. W. Jones, and F. C. Kafatos, *Biochim. Biophys. Acta* **700,** 42 (1982).

[11] T. I. Bonner, D. J. Brenner, B. R. Neufeld, and R. J. Britten, *J. Mol. Biol.* **81,** 123 (1973).

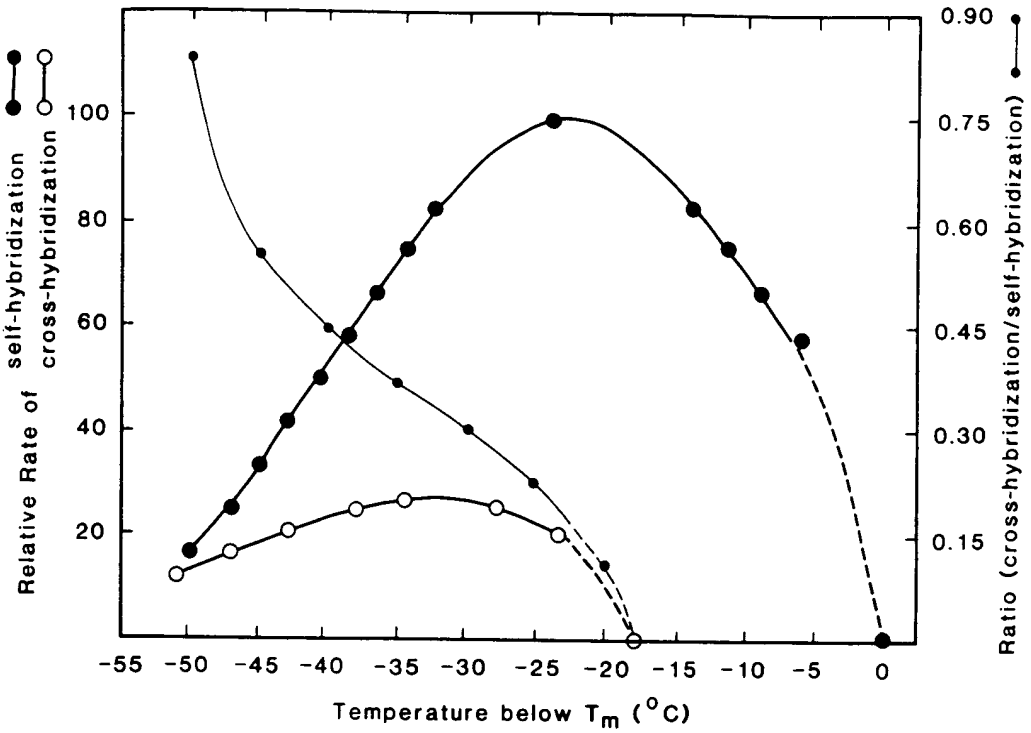


FIG. 1. Rate of reassociation as a function of temperature for normal (self-hybridization) and mismatched DNA (cross-hybridization). Data are replotted from Fig. 3 of Bonner *et al.*[11]; they are derived from normal T4 DNA and T4 DNA partially deaminated with nitrous acid. The large open and filled circles are from the best-fit curve plotted in the original figure, and dotted lines are extrapolations assuming that at T_m the rate of reassociation is zero. Under the conditions of the experiment (0.12 *M* phosphate buffer), the T_m of normal T4 DNA is 81°. The discrimination ratio, i.e., k of cross-hybridization divided by k of self-hybridization, is also presented.

curve of Fig. 1, for the reassociation of perfectly matched strands in solution.[11]

The problem of discrimination among related nucleic acid sequences is one of devising hybridization conditions that are stringent for some sequences and permissive for others. Although it is known that mismatching depresses both T_m and k, its effect on the relationship between k and criterion has not been studied in detail. Figure 1 presents one of the few sets of systematic data available for hybridization between normal T4 DNA and T4 DNA partially deaminated with nitrous acid (resulting in purine–pyrimidine mismatches that lower the T_m value by 18°). The temperature dependence of k for this "cross-hybridization" (T4 DNA vs deaminated T4 DNA) shows some similarity to that for "self-hybridization" (T4 DNA vs T4 DNA), but the rate constant is lower and reaches its optimal value at a lower temperature.

From data such as those of Fig. 1, one may calculate what we will call the "distribution ratio" for each temperature, i.e., the ratio between the rate constants for cross-hybridization and for self-hybridization reactions. The higher the value of this ratio, the easier it is to detect even distant homologs; the lower the ratio, the easier it is to discriminate between self-hybrids and cross-hybrids. It can be seen that the ratio varies with criterion in a sigmoidal manner. Although the exact shape of the sigmoid is not based on extensive empirical data, its general shape is not affected by minor variations in the individual k vs temperature curves: it depends on the bell shape of the curves and the shift of the cross-hybridization curve to lower temperatures.

The sigmoidal curve for discrimination ratio vs temperature has important implications for the study of multigene families. If we imagine a set of such curves, corresponding to different degrees of mismatching, we can understand why it is not possible to find a compromise criterion that would permit both detection of distant homologs and discrimination between close homologs. If both goals are to be attained, the following two-step procedure is necessary.

Step 1. To recover all members of a multigene family, an initial screen should be performed at a very permissive criterion. From Fig. 1, it can be seen that a criterion 40° to 50° below T_m is desirable if very distant homologs are to be recovered: at these temperatures the rate of self-hybridization is reduced two- to six-fold, but even distant homologs give a relatively strong signal. A practical limit is imposed by background problems. For example, if the probes have been cloned by G : C tailing, nonspecific hybridization due to the tails becomes significant at very permissive criteria.

Step 2. To discriminate between homologs, a very stringent criterion should be used in a second hybridization step. Since it appears that the discrimination ratio drops steeply near the T_m of the mismatched hybrids (Fig. 1), we recommend the use of criteria near the T_m of the relevant cross-hybrids. Temperatures as high as 5° below the T_m of the self-hybrids can be used without difficulty.

In theory, these two steps might be compressed into a single two-step experiment: After an initial hybridization under permissive conditions, poorly matched hybrids might be recognized by their release in a subsequent wash under stringent conditions. In practice, although this protocol is useful for discriminating between distant homologs,[3] we have found it not to be useful when the homologs are closely related. In part, the problem is technical. Although probes are for the most part reversibly hybridized, in filter hybridization a variable portion of the probe may become irreversibly bound. This problem can be minimized by scrupu-

lously avoiding drying the filter until after the melting reaction, but the precaution is not always sufficient.

A more important consideration affecting the design of the experiment is that a stringent hybridization and a stringent wash are not entirely equivalent. *A priori* the stability of a hybrid during washing depends on the T_m of the best-matched region of duplex. By contrast k reflects the nucleation frequency, which is an unknown function of the distribution of matched and unmatched stretches. In practice we find that a stringent hybridization is more discriminatory than an equally stringent wash.

Quantitative Considerations

It is clear from the preceding discussion that in the study of multigene families, selection of appropriate conditions requires attention to the parameters that govern T_m and k. We emphasize that systematic empirical data for matrix-supported reactions, and for mismatched sequences, are extremely limited. Thus, the appropriate conditions must be determined empirically in each case. However a starting point can be selected by calculations, and we summarize here the pertinent information from the literature for convenience in doing so.

The rate constant for renaturation of randomly sheared DNA in solution is given by the Wetmur–Davidson equation[12]

$$k = (k_N L^{0.5})/N$$

where L is the mean length of the reassociated duplex per nucleation; N is the complexity of the DNA; and k_N is a length-independent nucleation rate constant. The length factor is quite uncertain, because it attempts to incorporate effects of length on both nucleation frequency and yield of duplex per nucleation. Therefore, the equation can be used only for a rough estimate if the hybridizing strands differ in length. The constant k_N is influenced by such environmental factors as the incubation temperature, the salt concentration, organic solvents, and various polymers.[13] In an aqueous solution at an incubation temperature of $T_m - 25°$ and 1 M Na^+, $k_N = 3 \times 10^5$ liters/mole-second. Attachment of one of the reacting species to a matrix decreases k_N, probably by an order of magnitude.[13a] In addition, if a large amount of DNA is filter-bound, the hybridization may be limited by diffusion of the probe to the filter, unless the reaction vessel is shaken.[13a]

The effect of individual parameters on k is as follows. The change in

[12] J. G. Wetmur and N. Davidson, *J. Mol. Biol.* **31,** 349 (1968).
[13] R. Wieder and J. G. Wetmur, *Biopolymers* **20,** 1537 (1981).
[13a] R. A. Flavell, E. J. Birfelder, J. P. Sanders, and P. Borst, *Eur. J. Biochem.* **47,** 535 (1974).

optimal k at $T_m - 25°$ as a function of salt has been tabulated,[12,14] and is most dramatic below 0.4 M Na^+. An 1% increase in G + C increases[12] optimal k by a factor of approximately 0.018. Mismatching is reported[11,14] to depress the optimal k by a factor of 2 per 10° reduction in T_m. Concentrations of 80% formamide are thought to depress k by a factor of 3 for DNA–DNA hybrids and by a factor of twelve for RNA–DNA hybrids, although the data are limited.[15] Volume-excluding inert polymers can be used to increase k.[13,15a] In an aqueous solution of 1 M NaCl at 70°, the renaturation rate increases 100-fold when dextran sulfate is added to 40%.[13]

Concerning T_m, a large number of quantitative studies can be summarized as follows[11,15,16]: For a perfectly matched DNA duplex,

$$T_m = 81.5 + 0.41\,(G + C) + 16.6 \log(Na^+) - 0.63(\%\ \text{formamide}) - \frac{300 + 2000(Na^+)}{d}$$

where G + C = percentage of guanine + cytosine; Na^+ = molarity of (Na^+) or equivalent monovalent cation; and d = the length of the hybridized duplex in nucleotides.

It is prudent to bear in mind the limits of each term in this formula. The dependence on G + C is accurate over the range 30–75% G + C.[17] The salt dependence[16] is valid for 0.01 to 0.40 M (Na^+), only approximately so at higher (Na^+); T_m is maximal at about 1.0–2.0 M (Na^+). The depression of T_m by formamide[15] is greater for poly(dA : dT) (0.75°/1% formamide) than for poly(dG : dC) (0.50°/1% formamide), and the value used in the formula was derived from human rDNA (58–67% G + C). The effect of polynucleotide length on T_m is somewhat controversial, but the correction listed[14] seems valid over the range 0.05–0.5 M salt. Finally, the formula applies to the "reversible" T_m, e.g., as assayed by optical measurements. The "irreversible" T_m, which is of relevance for autoradiographic detection, is usually higher by 7–10° in aqueous solutions.[18]

For estimating the T_m of RNA–DNA hybrids, the paper of Casey and Davidson[15] should be consulted. The effect of formamide on such hybrids is nonlinear (unlike the effect on DNA–DNA hybrids), leading to the well-known preferential formation of RNA–DNA hybrids at high formamide concentrations. At 80% formamide, 0.3 M Na^+, RNA–DNA hybrids are

[14] R. J. Britten, D. E. Graham, and B. R. Neufeld, this series, Vol. 29, p. 363.
[15] J. Casey and N. Davidson, *Nucleic Acids Res.* **4,** 1539 (1977).
[15a] G. M. Wall, M. Stern, and G. R. Stark, *Proc. Natl. Acad. Sci. U.S.A.* **76,** 3683 (1979).
[16] C. Schildkraut and S. Lifson, *Biopolymers* **3,** 195 (1965).
[17] J. Marmur and P. Doty, *J. Mol. Biol.* **5,** 109 (1962).
[18] K. Hamaguchi and E. P. Geiduschek, *J. Am. Chem. Soc.* **84,** 1329 (1962).

reported to be 20–30° more stable than DNA–DNA hybrids. At 50% formamide, 0.3 *M* Na^+, we find that the difference is approximately 11° in the case of β-globin sequences.[3]

The effect of mismatching on T_m is obviously very important for studies on multigene families. The usual simplification is that T_m decreases by 1° for every 1 ± 0.3% mismatch,[14] reckoned for DNA–DNA duplexes of 40% G + C with randomly placed mismatches.[14] In another study, percentage homologies for BK and SV40 DNA fragments determined by heteroduplex analysis gave the best fit to actual sequence data when 0.5°/1% mismatch was used.[18a] Obviously the distribution of mismatched bases will have a critical determining effect. Consider an extreme example: sequence A differs from sequence B by 50%. If the mismatch is clustered so that extended regions of A and B are identical is sequence, a high T_m will be observed; by contrast, if the mismatch is dispersed so that every second nucleotide differs, no hybridization will occur. The effect of sequence on the hybridization of short oligonucleotides with mismatches has been studied.[19,20] The extrapolation of these very interesting results to real-life problems will require some judgment and additional experimentation. Conditions that permit detection of homologies between various unequally diverged regions of sequenced viral genomes have been investigated.[21] In the chorion system, we have noted that sequences that are highly divergent overall often show a smaller T_m depression than expected.[22] As an example, the B family clones pc401 and pc408 differ by 37.2% overall, but only by 13.6% in the conserved central domain[9]; the change in T_m of the cross hybrids, as determined from dot blots,[22] is approximately 10°.

It must be stressed again that the formulas given above were derived for solution hybridization, and thus are only rough approximations for filter hybridization. For example, it would not be surprising if the effective length of filter-bound hybrids is reduced by steric constraints. In agreement with this possibility, we have the impression that for filter self-hybridizations k declines more rapidly than expected from Fig. 1, when the temperature is raised to within 15° to 5° of the experimentally determined T_m. The T_m itself may be slightly depressed for filter-bound hy-

[18a] R. C. Yang, A. Young, and R. Wu, *J. Virol.* **34,** 416 (1980).

[19] R. B. Wallace, M. Schold, M. J. Johnson, P. Bembek, and K. Itakura, *Nucleic Acids Res.* **9,** 3647 (1981).

[20] S. Gillam, K. Waterman, and M. Smith, *Nucleic Acids Res.* **2,** 625 (1975).

[21] P. M. Howley, M. A. Israel, M.-F. Law, and M. A. Martin, *J. Biol. Chem.* **254,** 4876 (1979).

[22] G. K. Sim, F. C. Kafatos, C. W. Jones, M. D. Koehler, A. Efstratiadis, and T. Maniatis, *Cell* **18,** 1303 (1979).

brids. For example, for β-globin DNA–DNA dot hybrids of 51% G + C, 580 bp at 0.3 *M* Na^+ and 50% formamide, we have observed[3] a T_m of 56°, whereas the formula presented above would predict 60.6° for reversible T_m, and therefore at least 67° for irreversible T_m.

Two commonly overlooked variables in differential hybridization, which affect discrimination among nucleic acid sequences, are the extent of the reaction and the mass ratio of the radioactive probe to the unlabeled, filter-bound nucleic acid. In general, the hybridization of the probe will show the following kinetics when two (or more) filter-bound sequences react in the same bag with a probe in solution

$$dC/dt = -kC^2 - k_f IC - k_i H_i C \quad (1)$$

where C is the concentration of single-stranded probe at time t; k is the reassociation rate constant for the probe sequence in solution; k_f is the reassociation rate constant between the probe and the identical sequence, I, bound to the filter; and k_i is the reassociation rate constant between the probe and the cross-hybridizing sequence H_i, bound to the filter.

The kinetics will differ significantly, depending on whether the filter-bound nucleic acid or the probe is in excess, and in the latter case whether or not the probe can itself reassociate. For simplicity, let us assume that all filter-bound sequences are equal in mass ($I = H_i$), and that $k_f = k$. (In fact $k_f < k$, but the exact magnitude of the difference is not known, and it does not change the results qualitatively.) Then, if the filter-bound nucleic acid is in vast excess, for example, in typical dot blots,[3] the hybridization reactions follow pseudo-first-order kinetics; when they go to completion, the fraction of the probe hybridized to sequence i becomes

$$k_i \Big/ \sum_{i=1}^{m} k_i \quad (2)$$

where m is the number of filter-bound sequences.

At any time during the reaction, the ratio of the amount of probe hybridized to sequence i relative to sequence j is given by k_i/k_j (Fig. 2); i.e., the ratio is invariant with respect to time. In this case, discrimination between self- and cross-hybrids is not affected by the extent of the reaction (Fig. 2), because all the filter-bound sequences continuously compete for the same, limiting probe. (Discrimination will be affected if different samples are hybridized separately, but that situation does not normally apply.)

If the radioactive probe is in excess over the filter-bound sequences (e.g., typical Northern blots,[5] genomic Southern blots,[4] and initial screens of genomic or cDNA libraries), discrimination will generally decline as

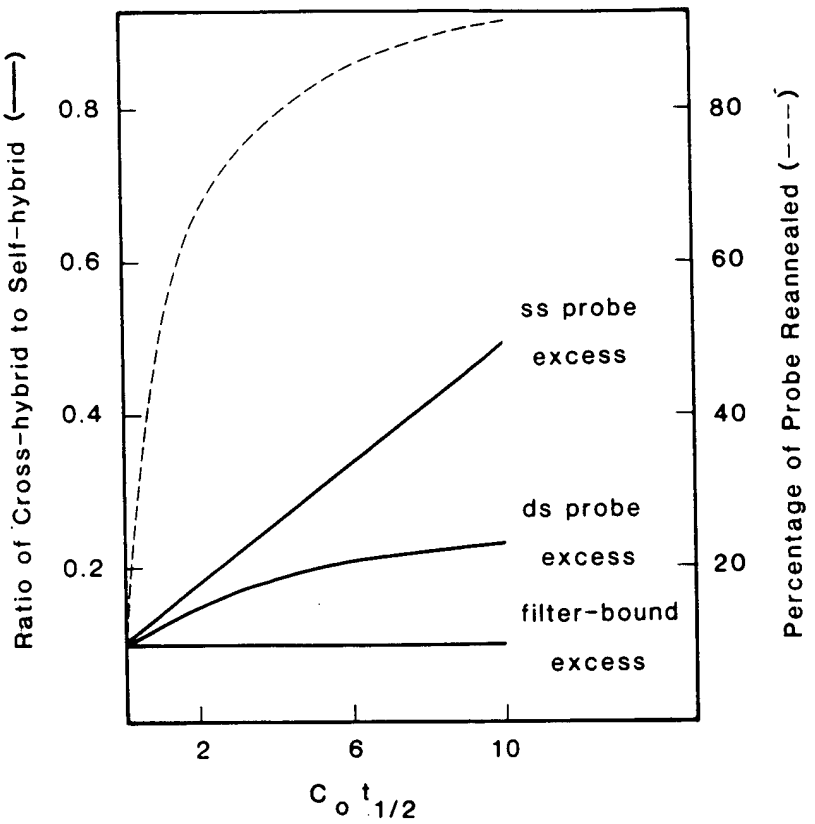


FIG. 2. Effective discrimination between self-hybrids and cross-hybrids, as a function of the extent of the reaction. The solid line shows the ratio of the amounts of probe hybridized to filter-bound heterologous and homologous sequences, $E_i(t)/E(t)$, for each of the three cases discussed in the text: filter-bound sequence in excess, double-stranded (ds), denatured probe in excess, and single-stranded (ss) (e.g., M13) probe in excess. The discrimination ratio, k_i/k, is assumed to be the same in all cases, 0.1. The dashed line shows normal C_0t kinetics of denatured, double-stranded DNA reannealing in solution.

the reaction proceeds. If the probe in solution can self-anneal, the relevant equation is

$$E_i(t)/E(t) = [1 - (1 + kC_0t)^{-n}]\,[1 - (1 + kC_0t)^{-1}]^{-1} \qquad (3)$$

where $E(t)$, $E_i(t)$ = extent of the reaction, i.e. the fraction of filter-bound DNA that has hybridized to the probe, for the self-hybrid and cross-hybrid i, respectively; C_0 = concentration of single-stranded probe at time 0; $n = k_i/k$, i.e., the discrimination ratio.

It can be seen that the effective discrimination, i.e., extent of cross-hybridization over self-hybridization, equals the discrimination ratio k_i/k very early in the reaction and deteriorates with the kinetics described above and as shown in Fig. 2.

If the probe in solution cannot self-anneal (e.g., M13 probes), the effective discrimination is given by

$$\frac{E_i(t)}{E(t)} = [1 - e^{-k_i C_0 t}]\,[1 - e^{-kC_0 t}]^{-1} \quad (4)$$

The parameters are defined above. Effective discrimination again equals k_i/k very early in the reaction, but deteriorates very rapidly (Fig. 2).

In summary, hybridizations should ideally be done for very short times: discrimination among various clones will then be maximal, and the same under conditions of either probe excess or filter-bound DNA excess. In practice, however, short hybridization times may not generate a sufficient autoradiographic signal. Thus, to distinguish among related sequences, it is best to use conditions of filter-bound DNA excess, as in typical dot blots[3] (see below).

A Strategy for Systematic Analysis of Multigene Families

Recovery of the Family

To begin the study of a multigene family for which a probe already exists, we recommend that a sublibrary be prepared by screening the appropriate library (cDNA or genomic) at a very permissive criterion [$k_i \simeq k$, $n = 1$ in Eq. (3)], using the Grunstein–Hogness[2,23] or Benton–Davis[1] procedures. At this stage we do not attempt to achieve discrimination, but only to recover as complete a collection of family members as possible; as discussed above, these two goals cannot be pursued simultaneously. A criterion of 30–35° below T_m and high salt (e.g., 0.6 *M* NaCl) for speeding up the hybridization reaction are recommended. To avoid unnecessary increase in the background, which is inherently high in the initial screening, the criterion should not be relaxed more than is necessary for the particular family. For completeness of the sublibrary, even very weakly hybridizing clones should be collected—especially since the plaque or colony size is frequently not uniform.

Such an initial screen will usually produce a mixture of true and false positives, which are distinguished by rescreening. Typical sublibraries that we have recovered in our study of chorion genes include from 200 to 600 clones, too many to make isolation of DNA from each clone practical. For sublibraries using bacterial plasmids as vectors, bacterial cultures are arranged in Microtiter plates, replica plated onto a nitrocellulose filter, placed on nutrient agar, and rescreened by the method of Grunstein and Hogness.[2] We have developed a similar replica plating method to rescreen

[23] M. Grunstein and J. Wallis, this series, Vol. 68, p. 379.

bacteriophage λ genomic sublibraries (see below). An example of the power and reproducibility of this method is shown in Fig. 3. Part a of Fig. 3 shows a rescreening of a genomic sublibrary under the same hybridization conditions as for the initial screen of the total library. This sublibrary had been selected to include genes closely homologous to the cDNA clone pc401, a B-family chorion sequence from *Antheraea polyphemus*.[9,22] APc110, a genomic clone containing two copies of chorion gene 401, and APc173, another genomic clone containing two copies of gene 10, a distantly related B-family sequence, were included as positive and negative controls.[9] A few blank spots were seen, corresponding to false positives of the initial screen, and a few clones grew poorly. Among the rest, some intensity differences were evident. However, even the negative control (APc173), showed significant hybridization under these conditions.

The next step is plaque or colony hybridization under stringent conditions. Figure 3b shows typical results, for the same sublibrary as in Fig. 3a. In this case, stringency was enhanced by a 15° increase in temperature and by use of a more specific probe. Instead of the complete pc401 DNA (565 bp), the probe contained a 257 bp fragment from the region corresponding to the 3′ end of pc401 mRNA. This probe encodes only a small portion of the conserved central protein domain, and consists largely of the more variable COOH-terminal arm sequence and nonconserved 3′ untranslated region.[9] Only 122 bp of this 3′-specific probe is well conserved within the B family (9% mismatch relative to pc10), whereas the entire pc401 probe includes a total of 255 bp of well conserved sequence. Clearly, the experiment of Fig. 3b achieved high discrimination. Hybridization with APc173 and with many of the clones in the sublibrary was essentially undetectable; only a few clones hybridized as intensely as APc110 itself, and others could be assigned to 3 or 4 classes on the basis of their degree of hybridization.

We perform plaque hybridizations to bacteriophage λ genomic sublibraries by the following procedure.

Preparation of Filters. Bacterial lawns are prepared by mixing 0.25 ml of 10 m*M* $MgCl_2$–10 m*M* $CaCl_2$, 0.5 ml of fresh bacterial overnight culture, and 6.5 ml of medium in 0.8% agarose. The mixture is poured over bottom agar (medium with 1.2% agar) in 150-mm petri dishes. Care should be taken to avoid air bubbles. We find that the bottom agar should be poured at least 2 days in advance to avoid condensation during phage growth. Allow the top agar mixture to gel for 10 min at room temperature. Using a replica plater,[24] transfer a small aliquot of each phage stock from

[24] M. Brenner, D. Tisdale, and W. F. Loomis, *Exp. Cell Res.* **90**, 249 (1975).

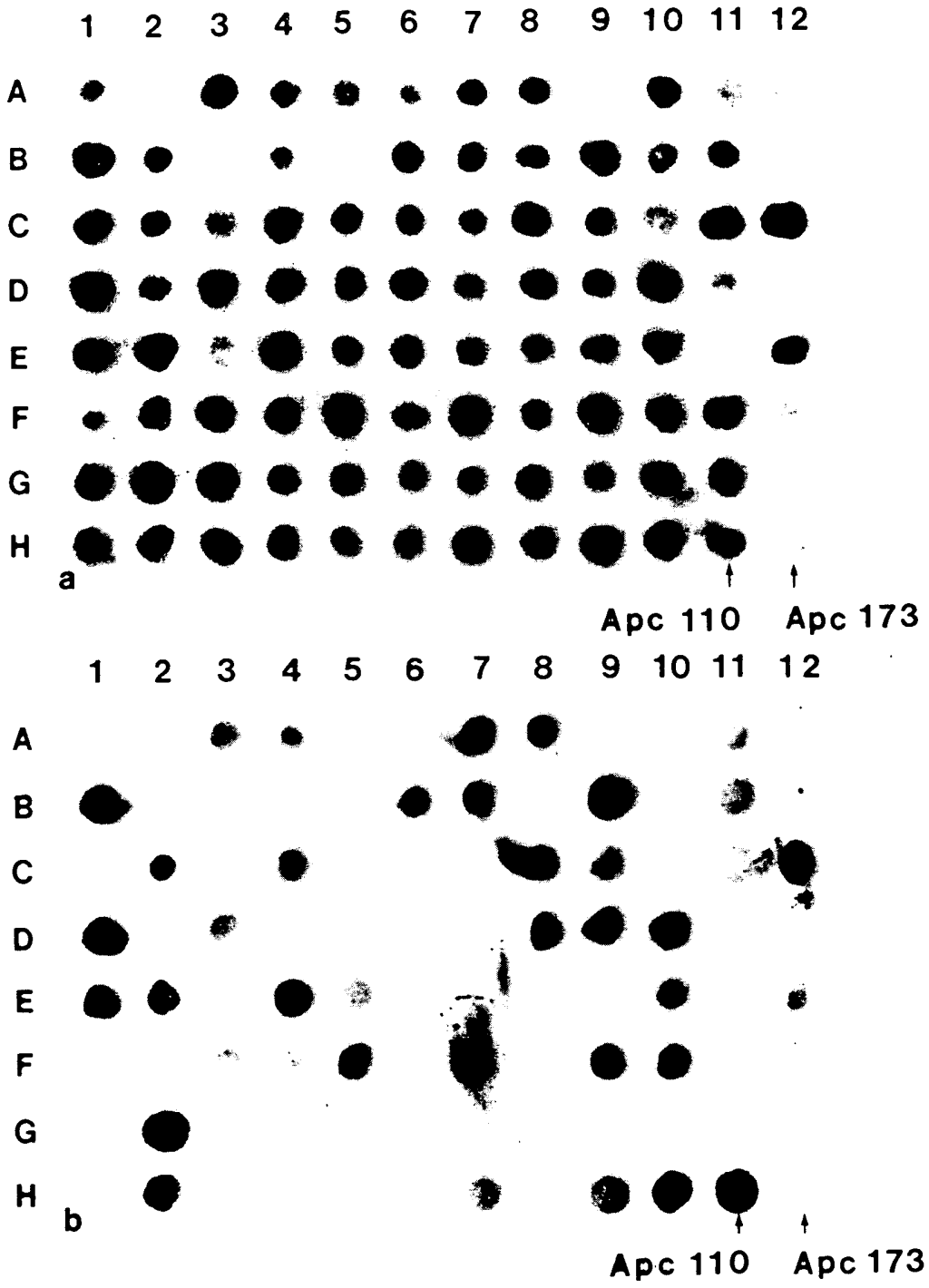


FIG. 3. An *Antheraea polyphemus* genomic DNA library was screened as described by Benton and Davis[1] with a combination of two chorion cDNA probes, pc18 and pc401. (The genes encoding chorion protein 18 are members of the A family and are each paired with a B-family gene encoding chorion protein 401.) Plaque-purified phage stocks from all positive clones were prepared. Aliquots of each stock were delivered into Microtiter wells and plated

Microtiter plates to the bacterial lawn. Allow the liquid to be absorbed into the agar (about 30 min), invert the plates, and incubate overnight at 37°. Uniform plaques of about 6 mm should develop. Chill the plates at 4° for 10–30 min before placing a sheet of dry nitrocellulose directly on the plaque-containing lawn. Precut nitrocellulose circles are commercially available, but we find it more economical and convenient to use rectangular sheets cut to 8.5 × 12.5 cm. Allow the DNA to transfer for 60 to 90 sec, remove the filter from the plate, and denature and fix the DNA to the nitrocellulose by soaking the filter in 0.1 *M* NaOH, 1.5 *M* NaCl for about 30 sec. Neutralize the filter by soaking in 0.5 *M* Tris-HCl, pH 7.5, 0.5 *M* NaCl for about 30 sec. Blot the excess buffer on 3 MM paper, air-dry, and bake under vacuum for 2 hr at 80°. We find that the same plate can be used to prepare at least four filters with no apparent loss in signal.

Prehybridization. Filters are prehybridized in 2–4× SET buffer (1× SET buffer is 0.15 *M* NaCl, 0.03 *M* Tris-HCl, pH 8, 1 m*M* EDTA), 10× Denhardt's solution, 0.2% SDS, and 50 μg/ml herring sperm DNA for at least 3 hr at the hybridization temperature.

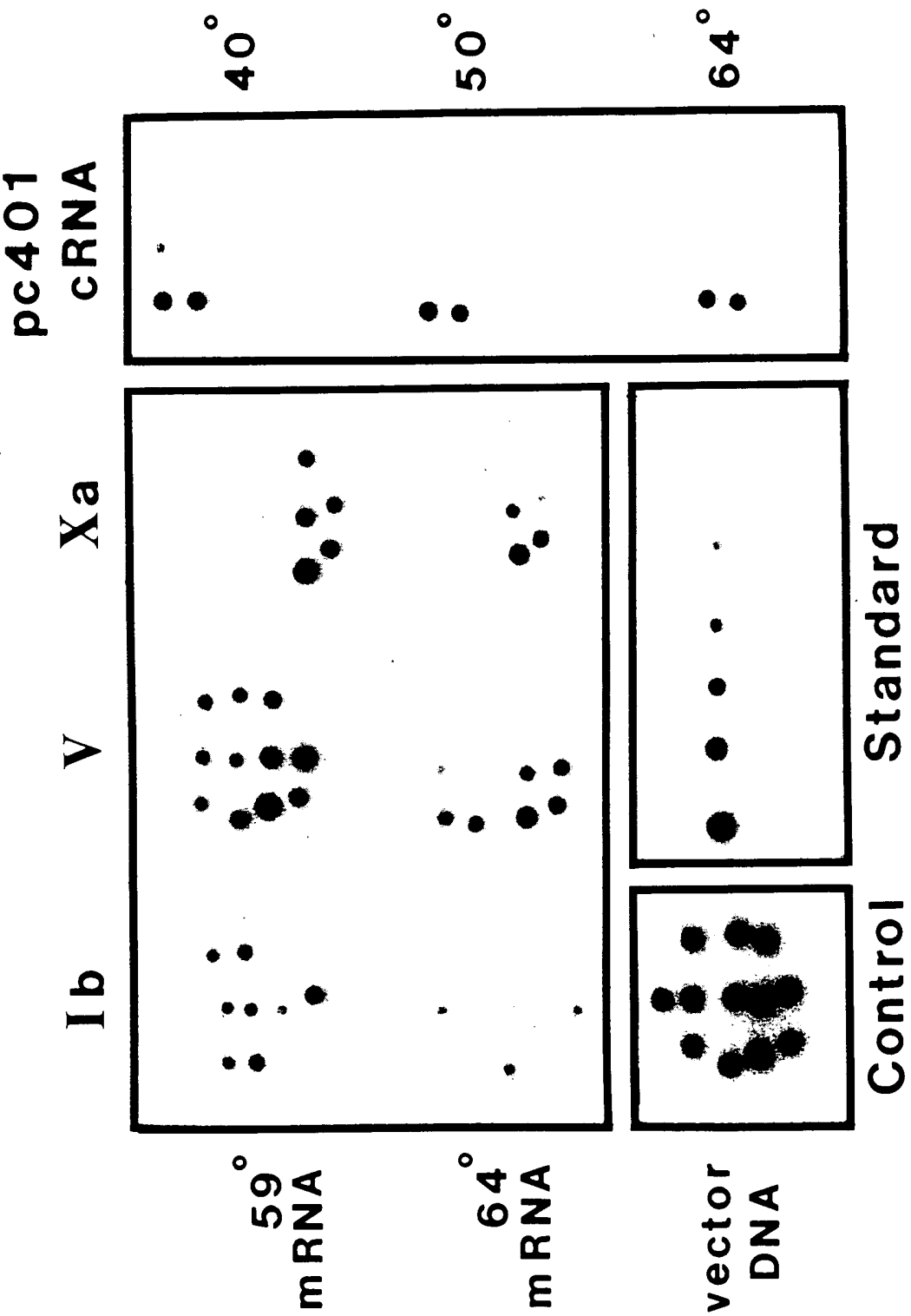
Hybridization and Washing. Filters are hybridized in prehybridization mixture supplemented with the radioactive probe and dextran sulfate to a final concentration of 10%. Although we normally hybridize overnight (12–16 hr), we find that the signals are sufficiently strong after hybridizing for only 5 hr. Filters are washed for 20 min each in two changes of 2× SET, 0.2% SDS and two changes of 1× SET, 0.1% SDS at the hybridization temperature.

Discrimination by Dot Hybridization

Once experiments such as those of Fig. 3b have focused attention on a limited number of clones, more refined analysis can be undertaken by dot-blot hybridization. The original paper describing the method[3] should be consulted, as well as a more recent review.[25] In this procedure, DNAs are purified from the clones of interest and immobilized on filters in precisely

[25] F. C. Kafatos, G. T. Thireos, C. W. Jones, S. G. Tsitilou, and K. Iatrou, *in* "Gene Amplification and Analysis," Vol. 2: "Structural Analysis of Nucleic Acids" (J. G. Chirikjian and T. P. Papas, eds.), p. 537. Elsevier/North-Holland, Amsterdam, 1981.

onto a lawn of bacteria. DNAs from the resulting plaques were transferred to nitrocellulose and hybridized to: (*a*) nick-translated pc401, at 65°, 0.6 *M* NaCl; (*b*) a nick-translated *Kpn*I fragment of pc401 representing the 3′ half of the mRNA, at 80°, 0.3 *M* NaCl. In each case hybridized filters were washed at the hybridization temperature in 0.3 *M* NaCl and 0.15 *M* NaCl. APc110 (a genomic clone containing two copies of 401) and APc173 (a genomic clone containing two copies of the distantly related, B-family chorion gene 10) were used as positive and negative controls.


40°
50°
64°
pc401
cRNA
Ib
V
Xa
Standard
Control
59°
mRNA
64°
mRNA
vector
DNA

equal amounts and in dots of uniform diameter. As a result, after hybridization the intensity of the various dots can be estimated semiquantitatively, by visual comparison to standards consisting of a dilution series of radioactive DNA directly spotted on a similar filter.

Figure 4 shows examples of dot-blot hybridizations.[22] The panel labeled "Standard" is a 2-fold dilution series that spans a 64-fold range of radioactivity. The panel on the right exemplifies the discrimination that can be achieved between homologs by judicious choice of conditions. The experiment was performed in 50% formamide, and pc401 cRNA was used as the probe. At 40°, all seven members of the B family included in this filter hybridized, albeit to different extents; five clones of the A family were also present in the filter, but did not detectably hybridize. At 50° and 64° two dots hybridized intensely and almost equally; one was pc401 itself (top) and the other pc602, which differs from pc401 by less than 1% mismatching. Weak hybridization with pc10 was observed at 50° (about 10-fold less than the self-hybrid), but none at 64°.

The remainder of Fig. 4 exemplifies a different, widely applicable use of dot-blot hybridization: assaying the concentrations of various sequences in a series of samples. Such assays are enormously more convenient and more sensitive than assays by liquid hybridization and, by careful attention to the conditions, are only slightly less accurate. The example shown permitted classification of 12 chorion cDNA clones into "early," "middle," and "late" developmental classes, depending on their preferential hybridization to mRNA preparations from the corresponding developmental stages (Ib, V, and Xa, respectively). In this case, the cloned DNAs were fixed to the filter and the labeled probes were mRNAs containing these sequences in unknown amounts. It is also possible to attach the unknown samples (either DNA or RNA) to the filter and probe them with a specific cloned sequence; examples of this approach are given by

Control Standard

FIG. 4. Discrimination of homologous sequences by dot hybridization.[22] *Top left:* Changing concentrations of specific chorion RNA sequences during development. Poly(A)-containing cytoplasmic RNA from *Antheraea polyphemus* follicles at stages Ib, V, or Xa was isolated, end-labeled with ^{32}P after alkali treatment, and hybridized to replicate filters containing equal amounts of spotted DNA from 12 different chorion cDNA clones. Hybridizations were performed in 50% formamide, 0.6 *M* NaCl for 28 hr (59°) or 40 hr (64°). As a control (*lower left*) one filter was hybridized with ^{32}P labeled pML-21 DNA, the vector used in cloning. *Right:* Cross-hybridization of chorion cDNAs. DNAs from 12 chorion cDNA clones (7 from the B family and 5 from the A family) were again spotted on nitrocellulose. ^{32}P-labeled 401 cRNA was prepared using *E. coli* RNA polymerase and the nuclease S1 excised insert of pc401. Hybridizations were performed in 50% formamide, 0.6 *M* NaCl at the temperatures indicated. *Standard:* A two-fold dilution series of [^{32}P]DNA spotted directly on nitrocellulose.

Weisbrod and Weintraub.[26] In the case of multigene families, it is important to distinguish the concentrations of identical vs related sequences. In Fig. 4, this was accomplished by performing the experiment at both 59° and 64°.

Exactly the same approach can be used to characterize members of the multigene families by hybrid-selected translation[6]: one hybridizes DNA dots under discriminating conditions with "probe" consisting of unlabeled mRNA and then melts the hybrids and translates the RNA in a cell-free system by standard procedures. If hybridization is performed under conditions that permit formation of some cross-hybrids, this technique can be used to detect the degree of homology between protein species for which sequence information is not available.[27,28]

An apparatus is now available from Bethesda Research Laboratories and Schleicher & Schuell for performing dot blots with multiple samples stored in Microtiter plates. Dot blots can also be performed manually, and we recommend the following procedure.[25]

Preparation of Filters. All filters are washed in water for 1 hr. A plain nitrocellulose filter (22 mm in diameter) is mounted on top of two nitrocellulose filters with grids, on a sintered-glass platform connected to a water aspirator. The filters are washed 3–4 times with 1 *M* ammonium acetate before spotting with DNA.

Plasmid DNA is linearized by restriction endonuclease treatment and digested with proteinase K (200 μg/ml in 50 m*M* Tris-HCl, pH 7.5; at 37°, 30 min each with 0.2% and 2% SDS). After phenol extraction the DNA is denatured in 0.3 *N* NaOH for 10 min and chilled; when needed, it is diluted with an equal volume of cold 2 *M* ammonium acetate to a concentration of 1.4 μg/ml. It is then taken up in a capillary pipette attached to a micropipette filler (Clay Adams suction apparatus No. 4555), and is spotted on the filter under light vacuum; once the pipette touches the filter, contact is maintained continuously, while the DNA solution is delivered slowly with a combination of vacuum suction and positive pressure from the filler. We routinely spot 0.7 μg of DNA in 50 μl per dot, using a 100-μl micropipette. The dot is washed with a drop of 1 *M* ammonium acetate, as is the area to be spotted next. After all the dots are made, the filter is washed again with 1 *M* ammonium acetate under suction, air dried, treated with 2× Denhardt's solution for 1 hr, drained, air-dried, and baked at 80° for 2 hr.

Prehybridization. The dried filters are wetted evenly by slow immersion in 10× Denhardt's–4× SET buffer and are shaken in that solution for

[26] S. Weisbrod and H. Weintraub, *Cell* **23,** 391 (1981).
[27] N. K. Moschonas, Ph.D. Thesis, Univ. of Athens, Athens, Greece, 1980.
[28] G. Thireos and F. C. Kafatos, *Dev. Biol.* **78,** 36 (1980).

at least 1 hr. They are transferred to sterile, siliconized scintillation vials containing blank hybridization mixture, e.g., 50% deionized formamide, 2× Denhardt's solution, 4× SET, 0.1% SDS, 100 μg of yeast tRNA and 125 μg of poly(A) per milliliter, and are incubated for at least 1 hr at the hybridization temperature.

Hybridization and Washing. The filters are hybridized for 16–48 hr in hybridization mixture (prehybridization mixture supplemented with radioactive probe). The temperature is selected to give the desired criterion of stringency. Washes are performed at the same temperature, twice each with 4×, 2×, 1×, 0.4×, and 0.2× SET, all with 0.1% SDS. Two final washes are performed at room temperature, with 0.1× SET without SDS. The filters are covered with Saran wrap and autoradiographed, either dry or moist (if melts are to be undertaken).

Discrimination by Southern and Northern Analysis

Discrimination between members of a multigene family, by careful control of hybridization conditions and selection of appropriately specific probes, can also be accomplished in Southern and Northern transfer experiments. While dot hybridization is preferable for quantitation and is somewhat more convenient, Northern analyses permit detection of the sizes of the hybridizing transcripts (which may vary during development), and Southern experiments using genomic DNA clones can identify restriction fragments that contain the cross-hybridizing sequences. Thus, these three filter hybridization procedures are complementary, and all are apt to be used in the study of a multigene family.

Figure 5 shows an example of discrimination by low- and high-stringency Southerns. The probe in this case was the high-cysteine cDNA clone m2574 of *Bombyx mori*,[29] and the filter-bound DNAs were derived from 18 overlapping genomic clones (B1 to B18) totaling 270 kb of continuous DNA spanning the high-cysteine region of the chorion locus.[30] Fifteen m2574-like genes were detected by low-stringency hybridization (75°, 0.6 *M* NaCl). High-stringency hybridization permitted identification of gene 12 as the one from which m2574 was derived, genes 14 and 15 as very close homologs, genes 8 and 10 as moderately close homologs, and the remaining 10 genes as more distant homologs. In this case, high stringency was accomplished by high temperature (85°, 0.3 *M* NaCl) and by the use of a relatively high concentration of DNase during nick-translation of the probe. The high DNase approach is recommended for maximizing

[29] K. Iatrou, S. G. Tsitilou, and F. C. Kafatos, *J. Mol. Biol.* **157**, 417 (1982).
[30] T. H. Eickbush and F. C. Kafatos, *Cell* **29**, 633 (1982).

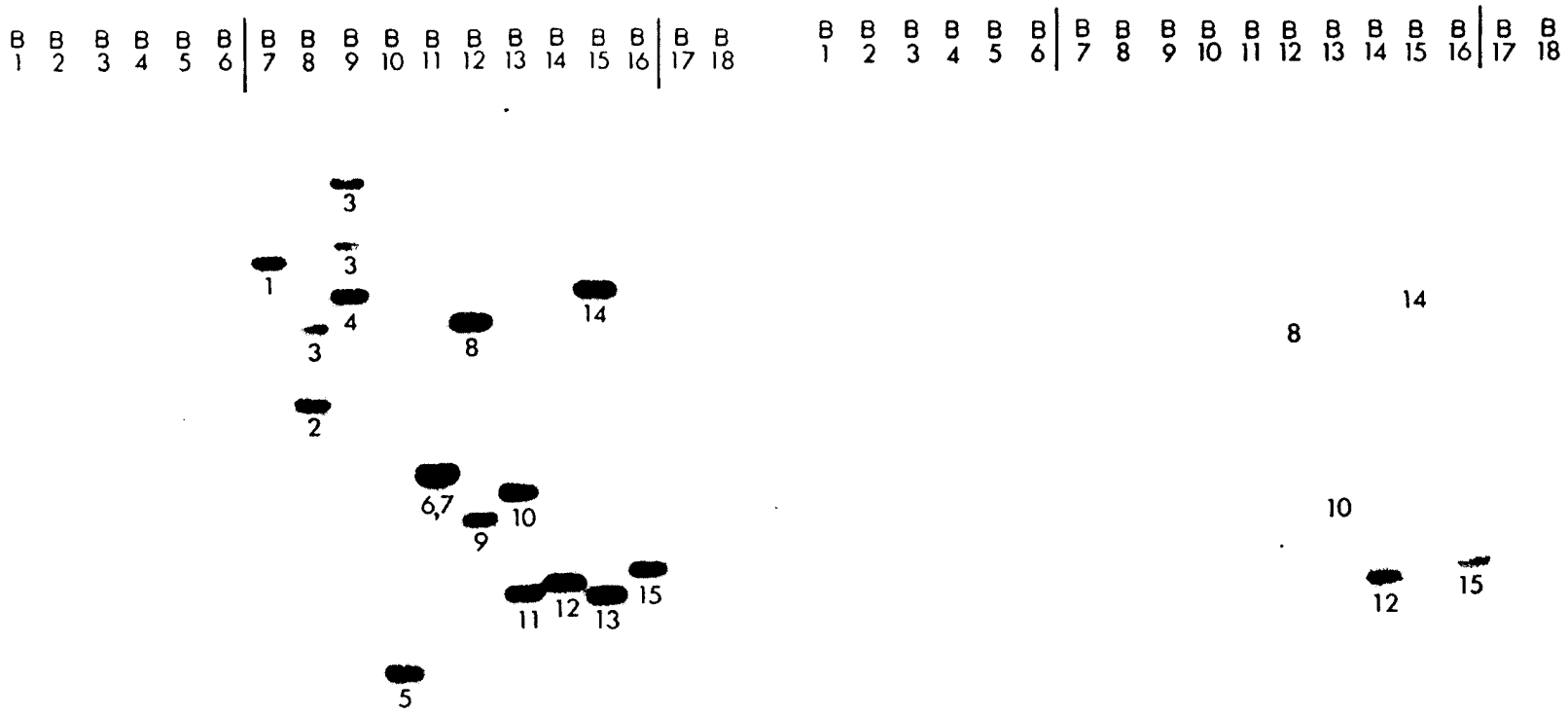


FIG. 5. Identification of a subset of related genes within a multigene family. Approximately 1-μg aliquots of DNA from genomic clones B1 through B18, which span the high-cysteine region of the chorion locus, were digested with *Eco*RI, electrophoresed through an agarose gel, transferred to nitrocellulose paper by the Southern method,[4] and hybridized with nick-translated cDNA probe m2574. The low-stringency hybridization (*left*) was at 75°, 0.6 *M* NaCl with the probe nick-translated in the presence of 5 ng of DNase I per milliliter. The high-stringency hybridization (*right*) was at 85°, 0.3 *M* NaCl, with the probe nick-translated in the presence of 125 ng of DNase I per milliliter. The numbers under each hybridized fragment refer to the m2574-like genes as numbered in Eickbush and Kafatos.[30]

discrimination when sequence information is not available, and therefore specific probes cannot be tailored as in Fig. 3b. High DNase results in short probe lengths, and, as a result, if mismatching is nonrandomly distributed, only the short fragments that might contain a conserved sequence will cross-hybridize at stringent conditions. By contrast, with long probes, signals from hybrids that are stable because of a short conserved sequence will be enhanced by the presence of covalently attached unpaired tails, resulting in decreased discrimination.

In conclusion, deliberate control of the stringency of hybridization is possible in all filter hybridization procedures and is a powerful tool in the isolation and characterization of multigene families.

Acknowledgments

Original work was supported by grants from NIH, NSF, and the American Cancer Society.

[20] Synthesis of ds-cDNA Involving Addition of dCMP Tails to Allow Cloning of 5′-Terminal mRNA Sequences

By HARTMUT LAND, MANUEL GREZ, HANSJÖRG HAUSER, WERNER LINDENMAIER, and GÜNTHER SCHÜTZ

Cloning of mRNA sequences after reverse transcription into cDNA copies[1-3] plays an important role in the analysis of gene structure and function. A variety of methods for cloning cDNAs in bacterial plasmids has been described.[4-10] In the method most commonly used,[7-9] the ability

[1] I. Verma, G. F. Temple, H. Fan, and D. Baltimore, *Nature (London) New Biol.* **235,** 163 (1972).
[2] D. L. Kacian, S. Spiegelman, A. Bank, M. Terada, S. Metafora, L. Dow, and P. A. Marks, *Nature (London) New Biol.* **235,** 167 (1972).
[3] J. Ross, H. Aviv, E. Scolnick, and P. Leder, *Proc. Natl. Acad. Sci. U.S.A.* **69,** 264 (1972).
[4] F. Rougeon, P. Kourilsky, and B. Mach, *Nucleic Acids Res.* **2,** 2365 (1975).
[5] T. H. Rabbits, *Nature (London)* **260,** 221 (1976).
[6] K. O. Wood and J. C. Lee, *Nucleic Acids Res.* **3,** 1961 (1976).
[7] T. Maniatis, S. G. Kee, A. Efstratiadis, and F. C. Kafatos, *Cell* **8,** 163 (1976).
[8] R. Higuchi, G. V. Paddock, R. Wall, and W. Salser, *Proc. Natl. Acad. Sci. U.S.A.* **73,** 3146 (1976).
[9] F. Rougeon and B. Mach, *Proc. Natl. Acad. Sci. U.S.A.* **73,** 3418 (1976).
[10] S. Zain, J. Sambrook, R. J. Roberts, W. Keller, M. Fried, and A. R. Dunn, *Cell* **16,** 851 (1979).

85°, 0.3 *M* NaCl, with the probe nick-translated in the presence of 125 ng of DNase I per milliliter. The numbers under each hybridized fragment refer to the m2574-like genes as numbered in Eickbush and Kafatos.[30]

ISBN 0-12-182000-9